An Integrated Approach to Achievement: Measuring the Development of Writing Skills in Kurdish Learners of English as a Foreign Language (EFL)

By

Haveen Muhamad Abdulmajeed

A Thesis Submitted to the University of Birmingham in Partial Fulfilment of the Requirements for the Degree of Doctor of Philosophy

School of English, Drama
and American & Canadian Studies
College of Arts and Law
The University of Birmingham
September 2016

Abstract

This thesis is a contribution to the field of learner corpus studies. It compares a number of different measures of accuracy and complexity in second language writing, applying those measures to a sample of 308 essays written by Kurdish university students majoring in English in three schools in Iraqi Kurdistan (at two years of study: third year and fourth year). It proposes an innovative method for measuring correctness, and integrates a number of different measures of accuracy into an Integrated Approach to Achievement. After the introductory first chapter, chapter two reviews previous research on learner corpora: their definition, design and compilation. It also describes the three constructs (Complexity, Accuracy and Fluency: or CAF) and critically reviews a number of the previous studies of CAF. Chapter three is a description of the process of data collection and the methods of analysis. Chapter four provides the results of applying traditional error analysis to a sample of the data collected and makes recommendations for measuring 'correctness' instead of concentrating on the analysis of errors. Chapter five operationalizes those recommendations, proposing an innovative method of assessment of accuracy in L2 writing by assessing 'correctness' as a replacement for the measurement of error using standard methods of analysis (the T-unit and clause-based correctness analysis). Chapter six also focuses on measuring correctness using a new method of analysis that takes various units into account and hence called the various-units-based correctness analysis. Chapter seven brings together all the measures of accuracy in a novel and integrated assessment method called an Integrated Approach to Achievement (IAA). The thesis then turns to measures of complexity. Various measures of syntactic complexity, including phrasal complexity, are described in chapter eight and applied to the data. Similarly, chapter nine uses a recently developed program called the Lexical Complexity Analyzer (LCA) to measure the lexical complexity of a subset of the essay sample. Chapter ten summarizes the research, discusses what the measures show about learner development in the sample essays, and makes some recommendations for how the research can be applied to pedagogy in the context of teaching English as a Foreign Language.

Dedication

- In memory of my parents and sister & her two little daughters

- To all my family with love

- To the souls of the martyrs of Peshmargas who have sacrificed their lives and to all the Peshmargas who are at the battle front-lines ready to sacrifice their lives for the sake of a stable and safe Kurdistan.

Acknowledgments

First of all, I thank Allah, the almighty for giving me the ability to pursue the degree of PhD and to perform a work such as writing a PhD thesis.

Furthermore, I am honored to have been supervised by Professor Susan Hunston. I am especially grateful to her for her invaluable comments, advice, and guidance that have enlightened me with significant knowledge without which this thesis would not have been completed. I am also indebted to the Ministry of Higher Education in Iraqi Kurdistan and particularly the Human Capacity Development Program (HCDP) for awarding me a scholarship to study for a PhD in the UK.

I would also like to thank Dr. Paul Thompson and Dr. Haiyang Ai for their kind support in teaching me how the tool of the Lexical Complexity Analyzer works. Moreover, special thanks also go to my colleague and friend Ahmeda Altoate, for his friendly and important cooperation in advising me to use the Dexter Coder for tagging and teaching me how it works. I would also like to acknowledge with great appreciation the support of my friends and colleagues at the English departments of the Universities of Dohuk and Zakho in the process of corpus compilation and design, especially, Suad, Rebeen, Hoger, Sharihan, Dilgash, Niroj, Shivan, Mahdi, Peshkafti, Baidaa, Vivian, Qasham, and Diman. I am also delighted to give a word of thanks to my other friends Shireen, Dina, Awaz and my flatmate, Awat for their great support. My students Awat, Areeba, Aisha, Khosheen, Jeen, Hamza and Omed deserve a lot of appreciation for providing me with a great support during the process of data collection. I should also acknowledge the work of proofreading this thesis by Dr. Sally Baggott that has really added to its value.

A word of thanks, also, go to the heads of English departments at the schools of Arts, Basic Education, and Languages at the Universities of Zakho and Dohuk; Dr, Izzuddin, Dr. Gharbi, and Dr. Aveen for allowing their students to participate as subjects in my study. Last but not least I would like to thank my family, particularly my brother, Ayas for giving me a special support in computer related issues.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

	Abbreviation	Stands for
1	A	Argumentative essay
2	1A3A	Subject no. 1 from school of Arts/year 3 who wrote an argumentative essay
3	1A3N	Subject no. 1 from school of Arts/year 3 who wrote a narrative essay
4	ADJV	Adjective variation
5	ADVV	adverb variation
6	1B3N	Subject no. 1 from school of Basic Education/year 3 who wrote a narrative essay
7	BNC	British National Corpus
8	CA	Contrastive Analysis
9	CAF	Complexity, accuracy and fluency
10	CCS	A completely correct sequence or structure
11	CCS/UT	The total number of completely correct structures divided by the total number of unanalyzable T units
12	CCT	A completely correct T unit
13	CCT/T	The number of completely correct T units divided by the total number of T units
14	CEA	Computer–aided error analysis
15	CIA	Contrastive Interlanguage Analysis
16	CL	Corpus Linguistics
17	CoA	Correctness analysis
18	CorrCN/W	Correct connectors per word
19	CorrPN/W	Correct pronouns per word
20	CorrART/CX	Correct article ratio
21	CorrDef/CX	Correct definite article ratio
22	CorrINDEF/CX	Correct indefinite article ratio
23	CorrPL/CX	Correct plural ratio
24	CS	Correct sequences
25	CSV files	Comma separated value files
26	CTTR	Corrected type/token Ratio
27	CVS1	Corrected verb sophistication-1
28	CVV1	Corrected verb variation-1
29	DC/T	Total number of dependent and embedded clauses/T units
30	1DE/T	First degree errors per T unit
31	2DE/T	Second degree errors per T unit
32	3DE/T	Third degree errors per T unit
33	EA	Error analysis
34	E/C	Errors per clause
35	EFT	Error free T Unit
36	EFC	Error free clause
37	EFC/C	The number of error free T clauses divided by the total number of clauses

38	EFCR	Error-free clause ratio
39	EFC/T	The number of error free clauses divided by the total number of T units
40	EFC/S	Error free clauses per sentence
41	EFL	English as a foreign language
42	EFS/S	Error free sentence ratio
43	EFT/S	Error free T-units per sentence
44	EFT/TC	The number of error free T clauses divided by the total number of clauses
45	EFT/TT	The number of error free T units divided by the total number of T units
46	EFTR	Error-free T-unit ratio
47	EFT/T	The number of error free T units divided by the total number of T units
48	EFT/TW	Error free clauses/ total number of words
49	EFT/W	Error free T units per word
50	ESL	English as a second language
51	Err/W	Errors per total number of words
52	E/T	Errors per T unit
53	FC	Fragment, correct
54	FDC	Finite dependent clauses
55	FI	Fragment, incorrect
56	FIC	Finite independent clauses
57	FL	Foreign language
58	FLT	Foreign language teaching
59	FS	Error in form spelling
60	GA	Error in grammar (article)
61	GADJS	Error in adjective, superlative
62	GDD	Error in grammar, demonstrative determiner
63	GDO	Error in grammar, determiner, possessive
64	GNC	Error in grammar,noun, case
65	GNN	Error in grammar, noun, number
66	GNP	Error in grammar (negative particle)
67	GPP	Error in grammar, personal pronoun
68	GPR	Wrong use of preposition
69	GPI	Error in grammar, pronoun, indefinite
70	GrE/W	Grammatical errors per word
71	GVA	Error in grammar, verb, aspect
72	GVAUX	Error in grammar, verb, auxilliary
73	GVM	Error in grammar, verb morphology
74	GVN	Error in grammar, verb, number
75	GVV	Error in grammar, verb, voice
76	GVT	Error in grammar, verb, tense
77	GWC	Error in grammar, word class
78	IAA	An Integrated Approach to Achievement
79	ICLE	International Corpus of Learner Language

80	L1	First language
81	L2	Second language
82	LCA	Lexical Complexity Analyzer
83	LCC	Error in lexis, conjunction, coordinator
84	LD	Lexical density
85	LexE/C	Lexical error per clause
86	LexE/LW	Lexical errors per lexical word
87	1L3N	Subject no. 1 from school of Languages/year 3 who wrote a narrative essay
88	LP	Wrong lexical phrase
89	LS	Error in single lexical item
90	LS1	Lexical sophistication-1
91	LS2	Lexical sophistication-2
92	LIWC	Linguistic Inquiry and Word Count
93	MFRM	Many-Facet Rasch Measurement
94	MLCS	Mean length of correct sequences
95	MODV	Modifier variation
96	MorE/C	Morphological errors per clause
97	MSTTR	Mean segmental type–token ratio
98	N	Narrative essay
99	NCCT	A not completely correct T unit
100	NCCT-P	A not completely correct T unit due to a sentence boundary missing and confusion punctuation error
101	NCCS	A Not completely correct sequence or structure
102	NCCT-P/T	The number of not completely correct T units due to punctuation divided by the total number of T units
103	NCCT/T	The number of not completely correct T units due to errors other than punctuation divided by the total number of T units
104	NCCS/UT	The total number of not completely correct structures divided by the total number of unanalyzable T units
105	NDW	Number of different words
106	NFC	Non-finite clauses
107	NF Clauses, Phrases/T	Non-finite clauses/phrases as nominal and adverbials / total number of T units
108	NF Clauses, Phrases/T	Non-finite clauses/phrases as nominal and adverbials / total number of T units
109	NFc/pN	Non-finite clauses and phrases as nominals
110	NFc/pA	Non-finite clauses and phrases as adverbials
111	NV	Noun variation
112	P	Participial phrase
113	PHI	Phrase Incomplete
114	PHU	Phrase Unclear
115	P/NP	Total number of participial clauses and phrases (*ing-participles and ed-participles*) as post nominal modifiers/ total number of noun phrases
116	PNM	Post nominal modifier
117	PP	Prepositional phrase

118	PP/NP	Total number of prepositional phrases as post nominal modifiers/total number of noun phrases
119	PU	Previously used in the chapters on accuracy
120	QC	Punctuation confusion
121	QM	Punctuation missing
122	QR	Punctuation, redundant
123	RC	Relative clause
124	RC/NP	Total number of relative clauses / total number of noun phrases
125	S()	Subject no.
126	SemE/E	Semantic errors per error
127	S/F language	Second or foreign language
128	SI	Sentence, incomplete
129	SLA	Second language acquisition
130	SL	Stretch, Ill-formed
131	SLW/LW	The total number of sophisticated lexical words divided by the total number of lexical words
132	SR	Sentence, Run on
133	SU	Sentence, unclear
134	SV(C)	Subject, verb, complement
135	SVO	Subject, verb, object
136	SVO(V)	Subject, verb, object, verb
137	SVT/V	Total number of sophisticated verb types divided by the total number of verbs
138	SWT/WT	The total number of sophisticated word types divided by the total number of word types
139	SynE/C	Syntactic errors per clause
140	TAALES	Tool of Automatic Analysis of LExical Sophistication
141	TC	Test code
142	T	T unit
143	To-inf or IF/NP	Total number of *to- infinitive* phrases as post nominal modifiers/ total number of noun phrases
144	To-inf. or IF	To-infinitve phrase
145	T/S	Total number of T units per total number of sentences
146	TTR	A type/token ratio
147	TWs	Total number of words in essay
148	UT	An unanalyzable T unit
149	UT/T	The number of unanalyzable T units divided by the total number of T units
150	VLexE/V	Verb lexical error per verb
151	WEFC/WC	Word in error free clauses ratio
152	Weighted EFTs	Weighted error free T units
153	WCR	Weighted clause ratio
154	WMS	Word missing single
155	WRS	Word redundant single
156	W/T	Words per T units
157	WT/√2W	The number of different word types divided by the square root of two times the total number of words

158	WT^2/W	Word types squared divided by the total words
159	XADJCO	Lexico-grammar, adjective with the wrong complementation
160	XADJPR	Lexico-grammar, adjective with the wrong preposition
161	XPRCO	A lexico-grammatical error of preposition with the wrong complementation
162	XVPR	Lexico-grammar, verb followed by the wrong preposition

Chapter One

Introduction

The purpose of this thesis is to measure complexity and accuracy in a corpus of essays written by Kurdish learners of English as a foreign language (EFL). These learners are university students at two universities in Iraqi Kurdistan. The motivation for carrying out this research on writing stemmed from the nature of the context of teaching and the challenges I faced as an EFL teacher, more specifically the teacher of English essay and research writing at one of the universities from where I took my data. Iraqi Kurdistan as an autonomous region in the north of Iraq is attempting to catch up with developments in the educational world. As English is a global language, the Kurdish people and authorities consider improving the learning of English as a foreign language as one of their priorities. The teaching of English, which used to start at the fifth stage of primary school, now starts at the first stage of the primary school. Also, the number of private schools with English as the medium of instruction has increased. However, people pay more attention to speaking and even use a spoken style in their written language. I could see this very clearly in the essays my students wrote. Another difficulty I faced in the teaching of writing was that students wrote essays that were conceptually dense but linguistically highly deviant. The essays could only be understood if taken in the wider non-linguistic context or as very small stretches of language. To first understand these essays and assess them was one of the very important challenges in my experience as a teacher. When I started to gather the data, I hoped to use it to investigate phraseological patterns, if the writing turned out to be of sufficiently high quality. However, the data very much resembled the type of essays I received from students during my teaching time (a sample will be introduced to the reader in this chapter). Hence, I decided instead to explore different methods for measuring accuracy and complexity in these essays in order to be able to come up with methods that best suit the analysis of such essays that are dense in meaning but contain a high number of errors.

In addition to the methods explored for evaluating complexity and accuracy, I have also used a new method of engaging teachers in the use of my methods for data analysis. 20% of the data analyzed for each chapter has been reanalyzed by a teacher in the university department in a discussion with me as the first rater. This was with the aim of evaluating

whether the proposed methods of analysis are usable by other teachers and ensuring that my data analysis has been carried out in accordance with the standardized rubrics that I have compiled and the ones that I have set. A further aim was to direct the attention of teachers in the department to the size of the problem in the written English of their students. The teachers in the department are the most relevant potential users of the method, and I wanted to see how enthusiastic they were to consider methods for devising new styles of teaching and assessment and to encourage them to use written rubrics for assessing writing. This is the reason why I have called this method 'user engagement'.

This first chapter describes the basic aims of this thesis, and lists the questions intended to be answered by this research. Furthermore, it provides an introductory overview of the methods of data analysis and the rationale for their use. It also discusses the problem that this thesis deals with and the proposed solution. In addition, an outline of the thesis is provided in the last section (1.5).

1.1 The main idea and topic

The purpose of this thesis, then, is to measure complexity and accuracy in the English written by Kurdish learners of English as a foreign language (EFL). These learners are university students majoring in English language and literature in three schools at two universities in Iraqi Kurdistan. The constructs to be measured are accuracy, and syntactic and lexical complexity in order to differentiate between the students who have performed well and those who have not as far as these constructs are concerned, and to see how the performance of third year students differs from that of fourth year students.

Complexity and accuracy are generally seen as parts of a triad, namely complexity, accuracy and fluency (CAF)[1]. Different measures have been proposed for their measurement as indicators in writing assessment and metrics of language proficiency and task performance. A great number of studies have used these measurement metrics (see Biber *et al*. 2011; Iwashita, 2006; Larsen-Freeman, 2006; Ishikawa, 1995; Homburg, 1984; Lim, 1984), and a number of studies have surveyed the measures in these studies and made comparisons between them (see Ortega, 2003; Wolf-Quintero *et al*, 1998; Polio, 1997). These metrics have been used as dependent variables to measure the effect of other factors (e.g. corrective feedback, planning, task complexity) on language performance. However, a number of researchers have used these metrics as independent

variables i.e. measuring them for their own sake and not for the sake of exploring the effect of other factors.

1.2 The main aims

The main aims of this thesis can be summarized in three points. The first is to objectively assess the English writing of Kurdish students at two levels of university study (third and fourth year) in order to effectively differentiate between their levels of performance. The second is to find new quantifiable, data-driven methods for measuring accuracy, and syntactic and lexical complexity as two facets of measuring language proficiency. The third is to evaluate language performance from a positive perspective (correctness) rather than the negative perspective (error).

1.3 The research questions

The research is intended to answer the following questions:

1. Is there evidence of improvement in performance between students in year three and four (in the three schools) in terms of accuracy and complexity?
2. How do different measures compare? What does each measure show?
3. Can we assess writing positively rather than negatively?

1.4 An outline of the problem and the proposed solution

I am an EFL teacher in the department of English at the University of Dohuk in Iraqi Kurdistan, where I collected a part of my data. Before starting my PhD, I taught essay writing to third year students for two years. I was thus interested to carry out research about writing, and initially I intended to look for phraseological patterns. However, two important points changed the direction of my research: (1) the main problems I have personally faced in teaching and assessing the writing of the students lie in the large quantity of errors and incomprehensible language; (2) the type of data I collected could not be analyzed in terms of phraseology, as the language was so deviant from the norm and hence could not be analyzed for issues such as collocations, lexical bundles and other topics in phraseology. These issues determined the aim for this thesis, which is to identify the best ways of analyzing the kind of essays that my students produced. More specifically the aim was to investigate different ways of assessing the accuracy and complexity in the essays of my students and students from other institutions as well.

I began by analyzing a small sample of the data to test the methods that could work well with the type of essays I have collected. The following is an example of the essays:

e.g. 1.1 S(211)

Do you agree/ disagree that the countryside life is much healthier than the city life?

I think is a good Idea life in the twon's is very much interesting that the city life and the twon's is a many different of the life of city and different people.
the people in the twon's are comfortable of every thing for example water, weather, the garden and the wide, eat, because in the twon's is amany farm and every thing is some vegetables, I think In the city no any thing is important of life because I agree of life the twon's .
I agree in the twons of that the city for example the weather is a very wonder and the climate degress changes is daily because in the summer is a very beautifull the farm and any beautiful and fun flower in the garden. I thing, I agree the twon's are do not ill the people but sometime's ill just afew, the people of the city many different because most of them is ill, therefore is arich, the weather Is durty in the city because a many office inside the city and the person is adifferent of the auther twon's, Realy I like live in the twon's and I do not like live in the city.
If the every person's is working for the parent's what the doing, will the known their parent's are wait for them, in the city or town's, therefore the city do not like the work just asleepor driving a car in the street, on playing football, go to the sinema and watch, but the town is more enjoy is wording who are doing in the farm or any aetherwork, the chidren of the twons are very understand of the life and is avery heading book and Is cleaver in the school I think agree is avery important of the twons of the anther city, the people in the city is increases the twons and the city are avery building, house, shop, maker but I do not agree life in the city give me a very happy and enjoy because the weather is avery fun and sweet, but the aether person like of live in the city but I do not agree with live of the city, also blessing I live in the town's is always, and I like a most of lives in the twon's.

Looking at examples such as this, it immediately became apparent that adopting measures of accuracy or complexity would be useful but also problematic. One such problem was that these methods tended to denigrate the students' abilities by prioritizing where they had failed over their level of success.

The first step I took was the application of error analysis. Although I applied a very comprehensive and detailed taxonomy of error analysis (the Université Catholique de Louvain's taxonomy), it was not straightforward to identify errors in the data as it contained more errors than correct forms. The analysis simply showed that the students produced many errors that hinder communication and which are almost impossible to identify and label. This led me to consider looking for correct forms rather than errors; I therefore devised a method that I named 'correctness analysis'. Researchers have used

this method before but only applied it to units like T-units[2] and clauses. They have measured the proportion of the error-free T-units and clauses. I tested this method, but it did not work well because the T-units and clauses are long units and they are very prone to errors. Moreover, the method did not take error severity into consideration so one simple error would render a whole T unit or clause wrong, and this simple error would be treated equally to T-units or clauses that contained severe errors. Consider the following two examples:

e.g. 1.2 S(211)

I think is a good Idea life in the twon's is very much interesting that the city life and the twon's is a many different of the life of city and different people. (very severe errors)

e.g. 1.3 S(13)

The first step on American land was happy and painful step, for she had to sit on the wheel-chair until she got out of the airport. (a few local errors)

Example 1.2 contains many errors that hinder communication. If one wants to correct it, then it needs to be rewritten. However, in example 1.3 there is only simple local errors like the lack of the use of the definite article *the* before the phrase *American land* and lack of the use of the indefinite article *a* before the phrase *happy and painful step*.

Because of my dissatisfaction with previous methods, I decided to devise a new method, which would identify any correct sequence be it a phrase or a clause or a sentence etc. However, even this method did not take error severity into consideration. I therefore progressed to uniting all the other methods and assessing the writings positively, giving this method the name of 'An Integrated Approach to Achievement' (IAA). As for complexity, in addition to measuring clausal and sentential complexity, I have also assessed the complexity of noun phrases in terms of post-nominal modifiers. This is because writing has been shown to contain longer and more complex noun phrases than spoken language (see Biber *et al*. 2011). The following chapters include a description, explanation and application of these methods.

1.5 The outline of the thesis

This thesis consists of ten chapters. Chapter two is a literature review. However, as well as chapter two, each chapter of data analysis contains a theoretical overview (i.e. a review of literature) of the concepts related to the analysis of that set of data. For example, the chapter on T unit and clause-based correctness analysis includes a review of the types of

units used in the analysis of written and spoken language and the chapters on syntactic and lexical complexity provide discussions about the way complexity is described, defined and operationalized in the related literature.

The outline of this thesis is as follows:

1. Chapter one provides an introduction to the topic and aims of the thesis. It also presents the research questions, the outline of the thesis and the possible ways of analyzing the data.
2. Chapter two presents the definition and design of learner corpora, and the type of comparisons made when carrying out research on them. The chapter also presents the background of complexity, accuracy and fluency (CAF) together with the ongoing debate about them. This is in addition to listing and explaining a number of the previous studies that have dealt with CAF and the measures they used to measure the components of CAF.
3. Chapter three explains the process of data collection and the methods of data analysis and describes the tools used for analyzing the data.
4. Chapter four presents the method of error analysis, its application to a sample of the data and the results this application yielded.
5. Chapter five introduces the method of correctness analysis using the T-units and clauses.
6. Chapter six introduces and applies the new method of 'various-units-based' correctness analysis to a sample of the data.
7. Chapter seven is an introduction, description and application of the new method of 'An Integrated Approach to Achievement' (IAA).
8. Chapter eight is an analysis of the data for grammatical complexity.
9. Chapter nine is an automated analysis of the data in terms of lexical complexity.
10. Chapter ten is a presentation of the conclusions, the pedagogical implications and the recommendations for future research.

Notes

1. See Skehan's model in chapter 2.
2. A T-unit is defined as an independent clause and its dependent clauses (Hunt, 1965). A detailed account of T-units and clauses is given in chapter 5 on correctness analysis.

Chapter Two

Learner Corpora and Complexity, Accuracy and Fluency (CAF): Background and Previous Research

2.1 Introduction

As this thesis is concerned with measuring accuracy and complexity in a learner corpus, this chapter will firstly review some basic information about learner corpora (definition, design and the principle of comparison), and secondly it will review the concepts of complexity, accuracy and fluency (CAF), the three dimensions commonly used to measure language proficiency, language performance and language development over time. This chapter answers a number of questions: (1) How are learner corpora defined in the literature? (2) How can a learner corpus be designed? (3) How is a comparison carried out in learner corpora? (4) What are CAF[1]? (5) Where did they emerge from as topics in applied linguistics and language acquisition? (i.e. their historical origins) (6) What research has been carried out on CAF? (7) What are the controversial points about CAF? Because these constructs are also dealt with in specific chapters on data analysis (chapters 5, 6, 7, 8, 9), a number of the related points (e.g. the units used in measuring CAF) will be left to be tackled in these chapters.

This chapter begins with a general discussion about learner corpora (their definition, design and the way they are compared to each other) in section 2.2. Section 2.3 presents the definition and background of CAF. Previous research on CAF is also described in section 2.4. Section 2.5 is devoted to the ongoing debate about CAF, and finally the chapter ends with some conclusions.

2.2 Learner corpora: definition, design, and the principle of comparison

A considerable amount of research has been carried out on learner corpora and how they can be designed and analyzed by computer tools. Most research has been carried out by Granger and her team (e.g. Paquot, 2013; Paquot, 2008; Granger, 2008, 2002, 1998a, 1998b; Meunier, 1998; Granger and Tyson, 1996). This is the team that compiled the influential International Corpus of Learner English (ICLE). Although this work is important and has inspired almost all the work done on learner corpora worldwide, there

is not sufficient space here to do justice to a full review of it; I will therefore restrict my comments to pointing out the aspects of the work which I make most use of in my thesis.

In order to define what a learner corpus is one needs to first give an overview of the way corpora are defined in the literature. For any linguistic data to be given the status of corpus, they have to share some important characteristics: they are expected to be textual pieces of language, authentic, representative of the variety of the language under investigation, electronically stored, compiled according to clear design criteria, and collected for the specific purpose of linguistic analysis (see the definitions provided by Hunston, 2002:2 and Sinclair, 2005 quoted in Cheng, 2012:30).

In a learner corpus, the texts are written by language learners. Granger (2002:7) bases her definition of learner corpora on Sinclair's (1996) definition of corpora and includes almost all the characteristics above. She defines it as follows:

> Computer learner corpora are electronic collections of authentic FL/SL textual data assembled according to explicit design criteria for a particular SLA/FLT purpose. They are encoded in a standardised and homogeneous way and documented as to their origin and provenance.

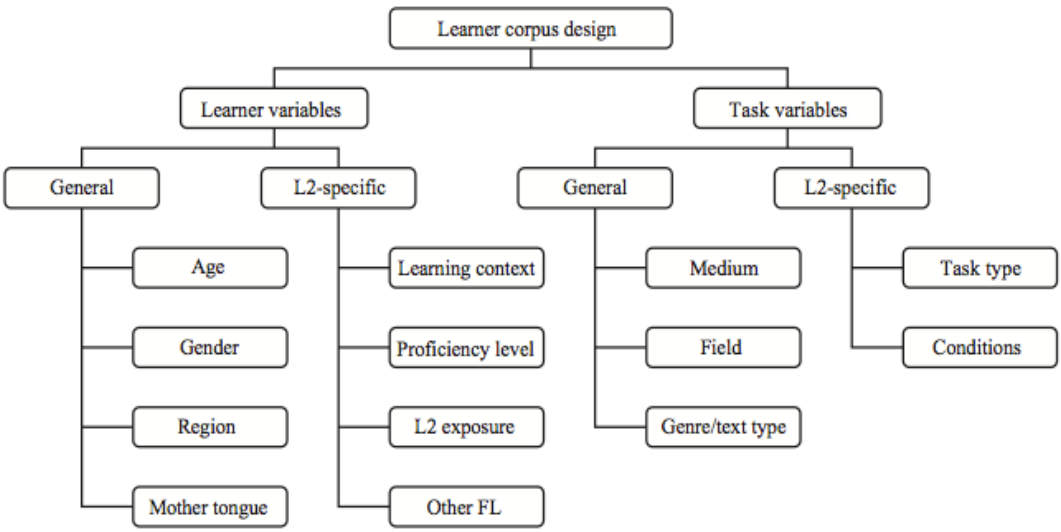
While Granger (ibid) has argued for flexibility as far as authenticity of data is concerned, she stresses the need for strict adherence to the textuality of data. Learner corpora should be comprised of continuous meaningful texts or parts of texts by foreign language speakers rather than sentences taken out of context.

One of the very distinctive features of corpora is the use of criteria to control their compilation and design; another is the care exercised in applying these criteria because the results of any corpus-based research "are only as good as the corpus" (Sinclair, 1991:9). However, these criteria are not the same in all cases but differ from one corpus to another depending on the type of corpus, the conditions under which the corpus is compiled and the purpose for which it is compiled and designed. Therefore, "a corpus cannot be judged except in the context of its purpose" (Hunston, 2008:155) as the context greatly impacts the choice of criteria.

Like any other type of corpus, a learner corpus is designed according to a set of criteria; this distinguishes it from the traditional collection of samples of learner language in cross-sectional studies, where little attention is paid to providing clear and complete

information about the learners and learning situation (Gass and Selinker, 2001 cited in Granger, 2008; see also Ellis, 1994). Granger (2008, 2002, 1998a) strongly argues for the importance of these criteria noting that they are what makes a learner corpus "systematic"[2] (Granger, 2008:263). This is because the nature of learner language is very variable due to the fact that it is affected by different "linguistic, situational and psycholinguistic" factors, and not controlling for such factors may make the results of learner corpus research less reliable (Granger, 2008:263) and not replicable. Granger (ibid: 264) presents a list of these variables in the following diagram.

Figure 2.1 learner corpus design



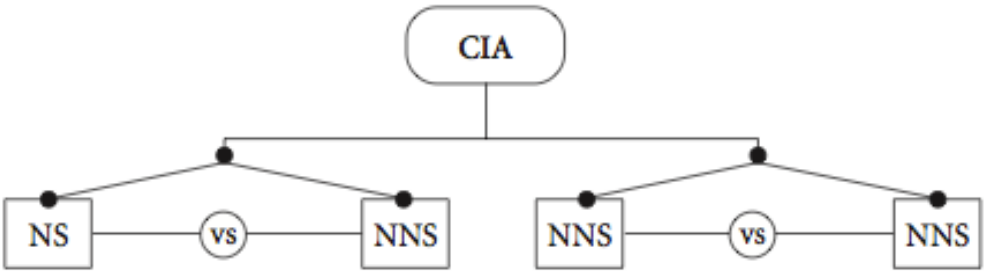
Following Ellis (1994), Granger divides the variables considered in learner corpus design into learner-specific and task-specific. Both these categories consist of (1) general criteria, which are applicable to all types of corpora and (2) L2/FL specific criteria. Thus, as is evident in figure 2.1, the learner specific criteria comprise age, gender, region, and mother tongue of the learner as general criteria and learning context (whether the target language is FL or SL), proficiency level, exposure to the target language, and knowledge of other foreign languages as L2/FL specific criteria.

Task-specific variables have also been reported to have an impact on learner language. It is very important to distinguish between spoken and written corpora (medium) as both differ to a great extent from each other and within each of these media, different genres (e.g. narrative vs. argumentative writing or spontaneous conversation vs. informal

interview) vary in terms of different aspects of language. The field (the topic) also greatly affects the choice of vocabulary (Granger, 1998a). The task type or the activity the learner is asked to carry out to produce the language (e.g. role-play, informal interview, spontaneous conversation, argumentative essay or letter writing, picture description, see Granger, 2008) has an impact on the language production. The term "conditions" is an umbrella term covering all issues that may affect learner's production of language whether spoken or written (e.g. time: timed vs. untimed activity, support: use/no-use of reference tools, nature of activity: exam/non-exam activity etc.) (Granger, 2008). In my study, some of these criteria, as will be evident in chapter three, section 3.5, have been taken to mean criteria because I have decided to use them as factors that should be controlled before compiling my corpus. However, some others are only intended as important information to be collected which I may resort to for the purpose of data analysis.

Granger (2002,1998a) has also proposed a completely new approach towards traditional contrastive analysis (CA)[3]. In traditional CA, researchers compare the second language of the learner with their first language to identify the areas of difference and hence difficulty for the learners and/or show the areas of interference (negative transfer) from their mother tongues (L1) (see R. Ellis & Barkhuizen, 2005). The new approach of CA is one of the important approaches that is devised with the emergence of ICLE. Granger (ibid) has called this new approach 'contrastive interlanguage analysis' (CIA). In this approach, two main types of comparison are carried out (see Hasselgård and Johansson, 2011 and Granger 1998a, 2002). The first one is to compare the learner language (L2) with the native language (L1), e.g. the written English of German learners of EFL with the written English of native speakers of English. The second type of comparison is to compare the L2 with L2. That is to say, compare the L2 of different groups of learners with different mother tongue backgrounds or the same L1 learners but with different age groups or proficiency levels (see figure 2.2). The approach I will be following is the second one, which is to compare groups of the Kurdish learners of English that are at different years of study.

Figure 2.2 the types of CIA (from Granger 2002:12)



Most of the work carried out by the Center of Corpus Linguistics in the Université Catholique de Louvain and other researchers on learner corpora prioritizes computer error analysis[4] and under and over representation of lexical and grammatical items (e.g. Chen, 2013; Bolton *et al.* 2002; Altenberg and Tapper, 1998; Granger and Tyson, 1996). The work also deals with phraseology (e.g. Dutra and Sardinha, 2013; Ädel and Erman, 2012; Granger and Meunier, 2008; Granger, 1998b; Howarth, 1998), the way acquisition of some grammatical items takes place (Housen, 2002) and the way phraseology or "formulaic language" should be intensified in pedagogy (Meunier, 2012). However, although initially I thought this approach might be the one that I would take, my own research has taken me in a rather different direction, and I will therefore turn my attention now to complexity, accuracy and fluency (CAF), which is the model that has influenced my own research.

2.3 Complexity, accuracy and fluency (CAF): definition and background

"What makes a second language (L2) learner a proficient language user? And how can L2 proficiency be most adequately (i.e. validly, reliably and feasibly) measured?" These are questions raised by Housen *et al.* (2012:1). Housen *et al.* (2012, see also Housen and Kuiken, 2009) declared that L2 proficiency and L2 performance are multidimensional with three dimensions: complexity, accuracy and fluency (CAF).

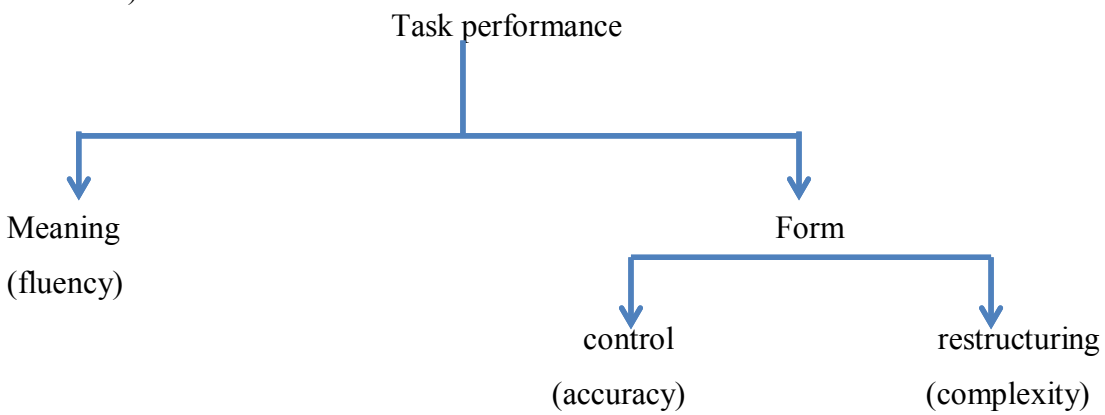
Two constructs of CAF, namely fluency and accuracy, have their origins in the 1980s (Housen and Kuiken, 2009). Housen and Kuiken (ibid: 461) have pointed out that it was Brumfit (1984) who first differentiated between "fluency-oriented" and "accuracy-oriented" activities with the former focusing on the "spontaneous" smooth use of language in speaking and the latter focusing on the controlled use of correctly structured language. Housen *et al.* (2012) date the origins of CAF to as far back as the 1970s, when

L2 researchers attempted to measure complexity and accuracy in second language performance, making use of the measures and units proposed by L1 researchers like Brown (1973) and Hunt (1965).

The origin of CAF is also explained in R. Ellis and Barkhuizen (2005). They argue that the structures that are learned "non-developmentally" (i.e. at any time) are subject to different ways of manipulation by the learner depending on socio-psychological factors such as the learner's orientation (see also R. Ellis, 1994). When the learners have a "segregative orientation", "they are likely to engage in a prolonged restrictive simplification" (R. Ellis and Barkhuizen, 2005: 140). They simplify the language and focus only on getting the meaning conveyed. These learners adhere to fluency at the expense of complexity and even accuracy. However, when the learners have "integrative orientation", they try to "complexify" their language by adhering to the form of the language rather than meaning, which means they want to speak a correct complex L2. With these learners, fluency is sacrificed for complexity and accuracy.

The three dimensions (i.e. CAF) were later united in one model in the 1990s by Skehan (1996, 1998 as cited in Housen *et al.* 2012 and R. Ellis and Barkhuizen, 2005). The model divided proficiency into *form* and *meaning* with fluency being meaning-oriented and accuracy and complexity being form-oriented. The following figure demonstrates this model:

Figure 2.3 Skehan's model for task performance (taken from R. Ellis and Barkhuizen, 2005:143)



Foster and Skehan (1996) also subsume complexity and accuracy under form and fluency under meaning. They relate complexity to the learner's risk taking in using elaborate language, defining it as a construct that "emphasizes the organization of what is said and draws attention to the progressively more elaborate language that may be used, as well as

a greater variety of syntactic patterning." (303) They define accuracy as "freedom from error" and fluency as "the primacy of meaning and the capacity to cope with real-time communication" (304) (see also the definitions provided by Housen *et al*. (2012), R. Ellis (2008), R. Ellis and Barkhuizen (2005), R. Ellis (2003) and Wolf-Quintero *et al.* (1998)).

2.4 Previous research on CAF[5]

Researchers have dealt with CAF in three different ways. First CAF were measured to indicate how corrective feedback, task complexity, pre-task planning, on-line planning[6] etc. impact second language acquisition in general and task performance in particular. Later, CAF became the focus of research as independent variables[7] (i.e. CAF were studied as the main issues, meaning that the research has investigated CAF for their own sake or has tackled issues that are only related to CAF, e.g. their measures, the basic units of their analysis, the issue of validity and reliability etc.). In addition, CAF were also researched in terms of psycholinguistics[8] or as Housen *et al.* (2012:2) call it "the primary epiphenomena of the psycholinguistic processes and mechanisms underlying the acquisition, representation and processing of L2 systems". This is concerned with the relation between CAF, the mental representations and processing of language and the way the knowledge about these representations is automatized. However, this type of research is not dealt with here because of space limitation and its irrelevance to this thesis.

2.4.1 CAF as the main focus

The studies listed and described in this sub section are chosen in order to (1) present a good introduction of CAF and their measures to the reader, (2) introduce the reader to the idea that CAF can be measured automatically by using computer tools and (3) describe a number of studies that have summarized a lot of the research carried out on CAF (in the form of research syntheses). For example, Polio (1997), Wolf-Quintero *et al.* (1998), and Polio and Shea (2014) are good examples of useful surveys in the field besides being independent studies of different measures of accuracy and complexity.

2.4.1.1 Polio (1997), Polio and Shea (2014) and Wolf-Quintero *et al.* (1998)[9]

Polio (1997), Polio and Shea (2014) and Wolf-Quintero *et al*. (1998) are three important collections of research studies about CAF. In both surveys Polio (1997) and Polio and

Shea (2014) have highlighted that (1) there was little or no use of clear rubrics for what constitutes a unit (i.e. T-unit, clause, sentence) and what *error-free* means and (2) measures of reliability were not reported sufficiently. It is important, nonetheless, to quote what Polio (1997:129) mentioned in this regard "this does not mean that the studies were poorly done or that the results are unreliable. However, providing more information helps other researchers anticipate problems when using similar methods". Polio (ibid) also recommended that researchers always provide the reliability scores even on a sample of the data used.

Taking this into consideration, no matter what the criteria used by researchers should be, it is always better for them to provide a clear list of criteria clarifying how they have divided the text into units and decided about the status of units as either erroneous or error-free. The absence of such a list may decrease the reliability of the method and may not allow for a replication of the study. It may also increase the possibility that the researcher's method of segmenting the text into units and identifying the error-free units is based on intuition rather than rigor.

In addition to being surveys of a number of studies, Polio (1997) and Polio and Shea (2014) have conducted their own studies. In her study, Polio (1997) provided a list of criteria[10] for the sake of implementing a reliable and valid division of the text into units, and Polio and Shea (2014) have also provided a list of rubrics for error codification and grading error severity (see table 2.1 below). Both studies have also reported the reliability and validity of their research.

Polio's (1997) research study aimed to compare three measures of linguistic accuracy, namely the holistic measure, the ratio of error-free units[11], and an error classification with the basic aim of identifying the problems that accompany the use of these measures and in order to develop a method suitable for measuring linguistic accuracy in "ESL students' essay revisions" (Polio *et al.* 1996 cited in Polio 1997: 103). Polio described the holistic measure as not being appropriate for homogenous samples. In terms of the other two measures, namely EFT/T and error counts, Polio's discussion implies that she recommends EFT/T.

Polio and Shea's (ibid: 11) study consisted of applying ten of the measures surveyed in their research to the Michigan State University (MSU) data with the aim of testing the reliability of the existing measures and showing which measure can best "capture development" over time or is more valid. The measures applied in their study included

the holistic measure, the measures that involved error-free units, number of errors, and number of specific error types and measures for error gravity. As mentioned earlier, with their measures, they provided very clear criteria for error coding and T-units ranking. The following are their rubrics for weighting T-units:

Table 2.1 Polio and Shea's (2014:25) guidelines for weighting errors and grading error-free units[12]

Score	Description
10	*No errors.*
7	*Minor errors that do not interfere with comprehension such as articles, SV agreement, prepositions, tense/aspect, or word choice*
3	*More severe errors in syntax or word choice or a large number of morphological errors. Intended meaning is still clear.*
0	*T-unit is incomprehensible or contains severe syntax errors.*

Polio and Shea also referred to the issue of validity and reported two ways of testing the validity of the accuracy measures. The first one is to apply another measure of accuracy and compare the results. However, since there is no "established" (ibid: 12) measure of accuracy, the low correlation cannot indicate which one of the measures is invalid. Another way is when measures are used with different groups, their results are expected to differ. However, "no difference" cannot be taken as evidence that the measure is invalid (Pallotti, 2009 cited in Polio and Shea, 2014).

Polio and Shea (ibid) calculated both validity and inter-rater reliability of the measures. In regard to the validity of the measures, they pointed out that the measures that were based on error identification (EFT/T, EFC/C, Err/W, and weighted EFTs) correlated highly with each other. Nevertheless, no measure "jumped out as the best" in terms of validity (22). This is why Polio and Shea have suggested that researchers assess measures before using them. The reliability of the holistic measure was higher than in Polio 1997. The reliability on the other two measures, namely error-free units and error counts were 0.88 and 0.89 respectively. They attribute the achievement of high reliability and the lessening of rater disagreement to the category they added to the guidelines of coding errors where the T-unit was graded 0 if it was not comprehensible or if it contained more than 5 errors. The reliability of coding specific errors was low, and the weighted T-unit ratio was not a more reliable measure than the measures of error-free units. They found this outcome quite unexpected.

Another important point that Polio and Shea (23) brought to attention is that no accuracy index can be described as "universal". For example, T-units might not be good for low-level students because the possibility of being rendered wrong is very high. However, there is more possibility for them to be right with low-level students than with advanced students if those low level students write short sentences.

Complexity seems to follow a developmental route, and it seems to be in a trade-off with accuracy. These were among the conclusions that Wolfe-Quintero *et al.* (1998) reached at after listing and describing a good number of studies together with the measures they used as metrics for CAF (see also Ortega, 2003 for a research synthesis). They have declared that (1) the development in the complexity of structures over time proceeds in the following way:

Fragments

Main clauses

coordinate clauses

adverbial clauses

adjectival and nominal clauses

adjectival, adverbial, and nominal verb phrases

(2) these structures develop "non linearly" in the form of omega Ω with an increase and then a decrease, (3) accuracy and complexity are in a trade-off relation, with the omega shape of complexity contrasting with the U shape of accuracy. In other words, "as the structure grows in use, it becomes less accurate; as the use of the structure moderates, it becomes more accurate" (74), (4) the following measures: dependent clauses per clause, mean length of clause, mean length of T-unit and clauses per T-unit are the most convenient measures as they have developed linearly according to the level of proficiency and are associated with the program level and holistic rating.

2.4.1.2 Casanave (1994)

Whether all developments in second language or foreign language can be measured by the use of T-units is an important question to answer. Casanave, (1994) in her study of the journal writing of a number of her intermediate level students over three semesters in Japan, found that the writing of the students changed but in various forms that could not all be accounted for by T-units. Actually, she concludes that the improvement in the "sense of fluency, maturity of topic, depth of treatment, honesty and genuineness of opinions, ability to take on a new perspective, and reflectiveness" (181) cannot be seen by only measuring changes in the language produced by groups of students. This unquantifiable facet for improvement, I believe, can only be researched in case the researcher is dealing with advanced students who have reached a level of accuracy that makes their sentences, T-units, clauses or any units they produce easy to understand or at least comprehensible enough to decide about the other factors.

The measures Casanave used were the percentage of error-free T-units and the length of error-free T-units for assessing accuracy, the number of clauses per T-unit and the percentage of complex T-units for assessing complexity, and the number of words per T-unit for measuring length.

Another interesting opinion of Casanave (ibid: 198) is that "we tend to focus on what is wrong and try to make it right, instead of focusing on what is right and encouraging more of it". I agree with this conclusion that we must look for what is correct and encourage it.

2.4.1.3 Larsen-Freeman (2006)

Although Larsen-Freeman (2006) found a linear increase in the four measures of language performance i.e. accuracy, grammatical complexity, fluency and lexical complexity, there was a great deal of variation between her subjects on the same measure and in the same subject on different measures. She also observed a greater change in accuracy than the other constructs, which she attributed to the fact that accuracy measures might be more sensitive to change. Her qualitative analysis revealed a lot of variability within the same learner at different times. Hence, she recommends that researchers look at the "messy little details" (ibid:613) of the performance of tasks of language development. We can "consider each performance a new—stable and predictable in part, but at the same time, variable, flexible, and dynamically adapted to fit the situation" (ibid: 615).

The study conducted by Larsen-Freeman (2006), was an "exploratory" study on five Chinese female learners of English. She gave each one the same task to perform every six weeks (4 times in 6 months). The task was writing a narrative about an event that had previously happened to the learner. The learner was then asked to tell the same written story orally three days after writing it. For the quantitative part of her analysis, she used different measures like average number of words per T-unit for measuring fluency, average number of clauses per T-unit for measuring grammatical complexity, the proportion of error-free T-units to all T-units for measuring accuracy and a sophisticated type-token ratio[13] for measuring lexical complexity. For the qualitative part of her study, she divided the texts into "idea units" and compared them (598).

2.4.1.4 Biber *et al.* (2011)

Biber *et al.* (2011:7) have challenged the idea that "longer units and more subordination reflect greater complexity" and maintained that it has not been empirically well proven that T-unit based measures and dependent clauses are suitable measures for measuring writing development. They (ibid: 10) further argued against the two "stereotypical" viewpoints: (1) operationalizing grammatical complexity through measuring the amount of clausal subordination is the "best" procedure; (2) academic writing is more complex than speech *because* it relies heavily on clausal subordination.

They have expressed their conflicting opinion as follows:

> Linguists who have studied the grammatical characteristics of spoken and written discourse provide a completely different perspective on complexity, arguing that a dense use of clausal subordination is not typical of advanced academic writing. In fact, clausal subordination is much more prevalent in conversational discourse than in academic writing. In contrast, linguistic analyses of written academic texts show that they are composed primarily of embedded noun phrases and prepositional phrases, with comparatively few embedded dependent clauses. (9)

To support their argument, Biber *et al.* have conducted a large-scale corpus-based "bottom-up" study to investigate the presence of 28 grammatical features (finite dependent clauses, non-finite dependent clauses, and dependent phrases that can function

as adverbial, complement, and noun-modifier which are then subdivided into 28 specific features) in academic writing as opposed to spoken texts. The results their study yielded are completely the opposite of what researchers have taken for granted for years. They have found that conversation has heavily relied on more clausal subordination and other features that have previously been thought of as characteristic of formal writing. Other features were explored to be characteristic of the written performance. "Put simply, the complexity of conversation is clausal, whereas the complexity of academic writing is phrasal" Biber *et al.* (2011:22)

Biber *et al.* argue that they still believe that formal academic writing is more complex than speaking because it has more phrasal complexity. In this way, they acknowledge that nominalization and adding more items to noun phrases is a higher level of complexity compared to the addition of more dependent clauses. They have also criticized the use of T-unit length as an appropriate descriptor of complexity. They give the following two examples as evidence:

e.g.2.1

Well, since he got so upset, I just didn't think we would want to wait for Tina to come back.

e.g.2.2

This may be part of the reason for the statistical link between schizophrenia and membership in the lower socioeconomic classes.

They highlight that, based on the T-unit length measure, the first sentence is equally complex to the second (20 words per T-unit) and based on the subordination measure the first sentence is more complex than the second. On the contrary of these two conclusions, they have declared that it is the second one that is more likely to have been produced by an advanced student in formal writing than in speech.

2.4.1.5 Evans *et al.* (2014)

Researchers have also paid attention to testing the validity of individual measures and have tried ways of showing exactly that the variance in the degrees of subjects can be surely attributed to the factors they are measuring. Evans *et al.* (2014) conducted an interesting study of the validity of one linguistic accuracy measure as a metric of linguistic improvement in L2 language writing, namely *weighted clause ratio* (WCR) and made a comparison between it and another two measures of linguistic accuracy: the error-

free T-unit ratio (EFTR) and the error-free clause ratio (EFCR). The WCR measure is "based on the adequacy of the writer's conveyed meaning" (Evans *et al*. 2014:33) and hence it seems that it measures adequacy, which Pallotti (2009) considers a different dimension from complexity, accuracy and fluency (see section 2.5).

The validity of the measure WCR is determined by its ability to show "how accurate a piece of writing is" and also to "effectively differentiate between separate accuracy levels across learners" (ibid: 37). Evans *et al.* (ibid: 39) evaluate the "criterion-related" validity for the WCR. The criterion-related validity can be defined as the degree to which a measure "correlates with an independent indicator of the same underlying ability" (O'Donnell *et al.* 2012 quoted in Evans *et al*. 2014:44). In this study, Evans *et al.*, as mentioned earlier, have chosen error-free T-units ratio and error-free clause ratio to correlate with this measure.

Evans *et al*. emphasized that the variance in the degrees of subjects might not only be due to those subjects' writing ability in terms of accuracy but also might be due to topic difference, or rating quality or task order. For example, a student might produce a more accurate piece of writing when they write about an easy or familiar topic than when they write about a difficult or unfamiliar topic. In order to account for all these factors, Evans *et al.* (38) have used the Many-Facet Rasch Measurement (MFRM), which "transforms the person's ability estimates and item's difficulty estimates into measures called logits (or log odds ratios) [that] are the natural logarithm for odds ratios of success and are convertible to and from probabilities". The outstanding quality of MFRM is that it measures the reliability of the factors involved in the calculation. When the reliability is close to one, this means that the variance is due to that factor and not something else. Accordingly, when the students' reliability is close to one, this indicates that it is the students' ability that causes the variance in the scores. They found that the rater reliability was 0.0–.02, which indicated that variance in the level of students is not due to the raters' assessment. However, they concluded that there should be a topic control in future research because the reliability score for topic was high (0.59) but this was ascribed to the research design rather than to the WCR itself.

In the data analysis Evans *et al.* declared that WCR is less practical that EFTR and EFCR because it took the raters more time to code the data. In terms of separating the subjects into levels or groups, the WCR was not more efficient than the EFTR but even less efficient than the EFCR; whether it is a valid measure or not Evans *et al.* have suggested

that the WCR is a valid measure for measuring linguistic accuracy, and it may even work better with low-level proficiency learners. Also, the three measures have correlated significantly with each other, and this is what made Evans *et al.* claim that they have all measured the same construct.

In the WCR measure, Evans *et al.* based their ranking of clauses on Wigglesworth and Foster's (2008)[14] rubrics for weighting their units. What is important is that they did not quantify the errors but only took meaning into consideration stating the following:

> If the researchers were not confused about the meaning, a clause was rated a 1. If they began to wonder about the meaning of a clause, it was rated a 2. If they found the meaning very difficult or impossible to determine, the clause was rated a 3. (Evans *et al*. 46)

2.4.1.6 Neary-Sundquist (2014)

The proficiency level seems to have an impact on the amount of subordination, coordination and phrasal complexity that the learners produce. To test this, Neary-Sundquist (2014) conducted a study on a German learner corpus consisting of the performance in four tasks (two speaking and two writing tasks) by students at four different levels of proficiency.

The main rationale of the study was a call for "agreement on what constitutes syntactic complexity and how it is measured" (214) in order "for research on the development of complexity to progress" (214). Neary-Sundquist used the ratio of coordinate clause to the total number of clauses to measure coordination and the ratio of subordinate clauses to the total number of clauses to measure subordination. For measuring phrasal complexity, Neary-Sundquist used the mean length of clause-based measure on the justifications made by Norris and Ortega (2009) that the increase in the mean length of clause can be due to the addition of more items as pre- and post-nominal modifiers.

The results of the study made by Neary-Sundquist showed that proficiency level had an impact on the amount of subordination the students produce because the students at level one and two used less subordination than those at level three, and those at level two also used less subordination than those at level four. So, there was an increase from two subordinate clauses per 100 clauses at levels one and two to nine subordinate clauses per 100 clauses at levels three and four, although at level four there was a little decrease. The

case with the coordination measure was similar to the subordination measure except for the fact that there was a high decrease of coordination from level three to level four. The results of mean length of clause were somehow different. The increase took place not from level one to level two but among the same students at level one (between different semesters). There was a high increase from semester one to two; then, the mean length of clause decreased slightly at level three before it increased again at level four.

2.4.1.7 Kyle and Crossley (2015)

Not only have researchers carried out manual analysis of complexity but they have also used automatic tools. For example, Kyle and Crossley (2015) have used an efficient computer tool for the analysis of lexical sophistication, which is one dimension of lexical complexity. The tool is the Automatic Analysis of Lexical Sophistication (TAALES) which is freely available on internet and it "calculates text scores for 135 classic and newly developed lexical indices related to word frequency, range, bigram and trigram frequency, academic language, and psycholinguistic word information" (757).

Because Kyle and Crossley aimed at testing the extent to which the results of the analysis using this tool will correlate with the holistic rating of words and speaking proficiency, they used two corpora that have already been rated holistically by experienced raters for their lexical and speaking proficiency: a corpus of unstructured free-writes that have been written by English language learners and native English speakers and a corpus of independent TOEFL speaking samples. They have proposed that the analysis with this tool accounted for 47.5% of the variance in lexical proficiency holistic scores and 48.7% of the variance in speaking proficiency holistic scores.

2.4.2 CAF in relation to other factors

CAF have also been measured as dependent variables to measure the effect of other factors on learners' performance on tasks. The factors included corrective feedback, planning, cultural intelligence, task type, etc.

Different researchers have investigated the effect of corrective feedback on the performance of the L2 and FL learners and thus they have measured accuracy, fluency and complexity (e.g. Salimi and Valizadeh, 2015; Shintani and R. Ellis, 2015; Akbarzadeh *et al*. 2014; Abdollahifam, 2014; Jokar and Soyoof, 2014; Jung, 2013;

Gholizade, 2013; Marzban and Arabahmadi, 2013; Shintani *et al*. 2013; Sato, 2012; Riazantseva, 2012; Farrokhi and Sattarpour, 2011; Chu, 2011; Evans *et al*. 2011; R. Ellis *et al*. 2008; Bitchener *et al*. 2005; Fazio, 2001; Frantzen, 1995). Bitchener *et al.* (2005:195) have measured the grammatical accuracy of "53 post-intermediate ESOL (migrant) learners" to find out about the effect of direct corrective feedback. They have used error identification with classification (i.e. they identified the errors in some specific aspects and classified them accordingly). They recorded errors in prepositions, simple past tense and articles, and they found that feedback highly contributed to a reduction in errors (i.e. increasing accuracy) in the use of past tense and articles but not prepositions. Akbarzadeh *et al.* (2014), measuring accuracy with the number of errors per T- unit and complexity with the number of content words per T-unit, have also found that oral interactive feedback had a positive impact on the accuracy and complexity of a number of Iranian intermediate EFL learners' performance. Marzban and Arabahmadi (2013) have investigated the effect of written corrective feedback on the writing of EFL learners. They used holistic scoring and the percentage of the correct instances of the structures "conditionals" and "wish statements" for measuring accuracy, the number of dependent clauses (produced in 30 minutes) per total number of clauses for measuring complexity and total number of structural units produced in 30 minutes for measuring fluency (1001). Their results have shown that feedback increased accuracy but there was no clear effect on complexity and fluency. Reading the results of these studies, it can be observed that feedback plays an important role in language improvement, especially in terms of accuracy.

The effect of planning on task performance has also been researched widely (e.g. Ghavamnia *et al.* 2013; Ahmadian, 2012; Abdi *et al.* 2012; Salimi *et al*. 2011; Ong and Zhang, 2010; Ahmadian and Tavakoli, 2010; Ojima, 2006; R.Ellis and Yuan, 2004; Yuan and R.Ellis, 2003; Foster and Skehan, 1999; Skehan and Foster, 1997; Foster and Skehan, 1996 see also Salimi *et al*. 2015; Ahmadian *et al.* 2015; Bamanger and Gashan, 2015; Bagheri and Hamrang, 2013; Rahimpour and Hosseini, 2010; Y. Wang, 2008 for the effect of planning and task complexity on CAF). These studies have used different measures for measuring CAF and have generated mixed results. For example, Foster and Skehan (1996) counted reformulations, replacements, false starts, repetitions, hesitation and pauses to measure fluency. For measuring complexity, they used the number of clauses/c-units[15] and forms of verbs in terms of tense, modality, voice and aspect. Accuracy was measured by counting the number of error-free clauses. They concluded that planning has more effect on narrative and decision-making tasks than on the personal

information exchange task. More interestingly, they have found that accuracy and complexity are in a clear trade-off relationship with the personal tasks being more accurate and less complex and the narrative tasks being the most complex and the least accurate. This indicates that the impact of planning differs with the type of task and that accuracy and complexity might not improve together but rather one at the expense of the other.

Ghavamnia *et al*. (2013) have also found that fluency and complexity increased with pre-task planning but accuracy increased with on-line planning. Ellis and Yuan (2004:59) have researched the effect of "pre-task planning", "unpressured on-line planning", and "no planning" on the CAF of 42 written narrative texts by Chinese EFL university learners. They used (1) syllables per minute and (2) number of dysfluencies for measuring fluency; (3) the ratio of clauses to T-units, (4) the total number of grammatical verb forms like different tense forms, different modality forms, and different voice forms, (5) Mean segmental type-token ratio (MSTTR)[16] for measuring complexity; (6) the percentage of error-free clauses, (7) the percentage of correct verb forms for measuring accuracy. The results of the study indicated that pre-task planning increased fluency and syntactic variety but unpressured on-line planning had an impact in the form of increased accuracy. However, no planning had a negative effect on the three dimensions (complexity, accuracy and fluency). One can conclude from the results of these studies that it is not only the type of task that affects the role of planning in improving the components of CAF, as mentioned earlier, but also the type of planning itself. For example, online planning increased accuracy but pre-task planning increased fluency and complexity.

The effect of other factors such as task complexity (e.g. Salimi *et al.* 2012; Kuiken and Vedder, 2008), task repetition (e.g. Sample and Michel, 2014), task type (e.g. Vesal *et al.* 2015), content-based language instruction (e.g. Tai, 2015), peer-mediated and individual writing (e.g. Soleimani *et al.* 2015), type of writing or task type (e.g. Kuhi *et al*. 2014, Rezazadeh *et al*. 2011) and cultural intelligence or knowledge about English culture (e.g. Ghonsooly and Shalchy, 2013) on CAF has also been investigated. These researchers have used different measures to assess CAF and their research has yielded different results. The measures used by these researchers included the ratio of error-free T-units (Vesal *et al*. 2015; Salimi *et al.* 2012), clauses/AS-unit, words/AS-unit (Sample and Michel, 2014), the proportion of error-free clauses (Tai, 2015; Soleimani *et al*. 2015), the total number of errors per total number of words (Tai, 2015) etc. for accuracy. As for

fluency, the measures used were the number of syllables per minutes taken to complete the task (Ghonsooly and Shalchy, 2013), the average number of words (Soleimani *et al.* 2015), the average number of T-units in each composition (Soleimani *et al.* 2015), and the length of the T-units measured in words per T-unit (Tai, 2015; Salimi *et al.* 2012), and filled pauses/minute (Sample and Michel, 2014). The complexity measures used included the average number of words, T-units and clauses per text (Kuhi *et al.* 2014), the total number of different grammatical verb forms (Ghonsooly and Shalchy, 2013), the ratio of clauses to T-units (Tai, 2015; Kuiken and Vedder, 2008), mean segmental type token ratio (Ghonsooly and Shalchy, 2013), the D measure[17] (Sample and Michel, 2014), the ratio of dependent clauses per T-unit (Tai, 2015; Kuiken and Vedder, 2008). As for the results, Tai (2015), for example, has found that content-based instruction had a positive effect on accuracy but not on complexity. Vesal *et al*'s (2015) results showed that in the introduction and discussion tasks, both accuracy and complexity increased but fluency decreased. On the monologue task, accuracy and complexity witnessed a decrease but also number of repetitions increased, meaning less fluency as well. Ghonsooly and Shalchy (2013) found that 15% variance in fluency could be accounted for by cultural intelligence but only 5% and 3% of variance could be accounted for in the case of complexity and accuracy respectively. They, hence, recommended a focus on cultural intelligence in teaching.

2.5 The debate about CAF

There has been a great deal of debate among scholars about CAF. The points of debate have included the importance of comparative results between groups even if they show similarity, the difference between accuracy and comprehensibility, the definition of CAF, the relation of CAF to cognition and psycholinguistics, the interconnectedness of the CAF components, and the operationalization and measurement of CAF. The following is a summary of that debate.

1. It is important to compare groups even if results are similar.

Pallotti (2009) pointed out that researchers always look for differences between the results when they apply the CAF measures, and they consider the measures to be valid only when they approach the differences between the results that those measures have

yielded. Pallotti (2009) argues that researchers should not only look for areas of difference but similarity as well. Pallotti (590-591) states:

> If after an experimental treatment two groups of subjects don't show any difference, then this is not a non-result, but a result just as interesting as their being different. Likewise, if a measure does not change over time, this does not make it a poor measure, but perhaps a measure pointing to a trait that does not actually vary.

Pallotti also highlights that even if the difference is significant between two measures out of many, it has to be dealt with carefully especially when the P value is near to 0.05.

2. The difference between accuracy and comprehensibility

Although accuracy is the least ambiguous construct within CAF, Pallotti argues that researchers have to be cautious when they measure it. For example, he points out that "accuracy" and "comprehensibility" are two different constructs and they should be distinguished when researchers measure them. He argues that a 100-word text that contains ten errors that do not hinder communication is not more accurate than a 100-word text that contains ten errors that hinder communication but it is only more comprehensible. In a sentence like *me no likes go dance*, the grammar of the target language is very considerably violated but this still does not make the sentence incomprehensible. However, a phrase like *colorless green ideas* is well-formed but "communicatively inadequate" (592).

In the data of the present study, the same problem occurred. In the students' essays, there are sentences that do not greatly violate the grammar of English but are still incomprehensible. There are also a number of units which are too ill-formed to be analyzed in terms of error analysis but are still understandable. Consider the following examples:

e.g. 2.3 S(11)

the first thing was very funniest for me is that when I became student in primary school because before I 'm going to school I saw my sister they are studing and they can write these all thing that make me to love school and became a student and the first day first day was so funniest for me because one of the small dream in my life it became true

e.g. 2.4 S(15)

So we have to remember every time when you laugh there are some thing can remove that thing

e.g. 2.5 S(47)

The way of living in country from the city life in countryside, we can not it is calm

The first stretch is very ill-formed but is still understandable. One can understand that the writer means that *the first thing that made me happy was my first day at primary school because before I attended school, I could see my sisters writing things. This made me love school and gave me the desire to become a student, so the first day was a happy day for me because my dream came true.* Example 2.4 could be considered as better than example 2.3 but it is unclear. It is not clear what the writer meant. Example 2.5, however, is both unclear and ill-formed.

So, although Evans *et al*. (2014:36) claim "as communicative adequacy deteriorates, it will always be associated with concomitant failures in accuracy", this might not be true in all cases and the reverse may not be true either. Although Evans *et al.*'s (2014) approach looks robust because it is dependent on a scale for rating accuracy in terms of the (in)comprehensibility of the unit of analysis, it does not account for units that are comprehensible but which very much violate the grammar of the target language.

3. Definition of CAF

Housen *et al.* (2012) have argued that CAF have been defined differently by different researchers, and that the definitions are sometimes vague. This has led to inconsistency of results in different studies and even within the same set of studies. They also argue that in most research CAF are treated as a single dimension rather than as a multidimensional construct.

Starting with accuracy, Housen *et al.* (2012) state that although it is not a difficult term to define, it is still problematic. The main concern, as they highlight, is related to the notion of errors as deviations from the norm and how they are defined. Housen *et al.* (2012) suggest that this "A" should stand for "appropriateness" and "acceptability" in addition to accuracy. Nevertheless, as mentioned earlier, Pallotti (2009) argues that acceptability and adequacy are two different constructs from CAF. This problem, I believe, can be solved if the researcher identifies the norm of English against which they are measuring the correctness of learner language (e.g. British English, American English or the

researcher's own English as the case with this thesis). Also, the researcher can define what can be considered an error. Is appropriateness and acceptability included in the definition of error? Or only accuracy?

Complexity has also been described as a very complex component of CAF as it is in itself multidimensional and multicomponential (Housen and Kuiken, 2009). Housen and Kuiken point out that complexity can be divided into cognitive complexity and linguistic complexity. Hence, when researchers measure complexity, they have to be aware of that and define what they actually mean in terms of the type of complexity they measure.

The criticism about fluency represented in the fact that although fluency is of three types (*break-down fluency* i.e. rate of pausing, *repair fluency* i.e. rate of repetitions and self-corrections, and *speed fluency* i.e. rate of delivery), it is only the "speech" rate and rate of dysfluency that are most commonly measured (Lambert and Kormos 2014, see also Housen *et al.* 2012 for a similar criticism).

4. CAF and its relation to cognition and psycholinguistics

Many writers (e.g. Ruiz-Funes, 2015; Sample and Michel, 2014; Salimi and Dadashpour, 2012; Housen *et al.* 2012; Kuiken and Vedder, 2008) have referred to two opposing cognitive models of CAF: the cognition hypothesis of Robinson and the trade-off hypothesis of Skehan (see Foster and Skehan, 1996, Skehan 1996 in relation to his model of language production described above). Robinson (2001, 2005) claims that both accuracy and complexity increase with the increase in task complexity. However, Skehan bases his model on the information processing theory, arguing that human attentional capacity is limited and humans cannot attend to different things at one time. Skehan further argues that a person can either attend to form or meaning (see the model above) and that focusing attention on one of them may be at the expense of the other. This has been the concern of a lot of research (e.g. Ruiz-Funes 2015; Sample and Michel, 2014; Salimi and Dadashpour, 2012; Kuiken and Vedder, 2008) which investigates the effect of other factors on CAF e.g. careful online planning, task complexity etc. The investigation of the effect of these factors is to try to explore which factor can help the learner attend to two or three dimensions of language production, namely complexity, accuracy and fluency.

5. The interconnectedness of the CAF components.

Both Housen and Kuiken (2009) and Housen *et al.* (2012) have pointed out that although complexity, accuracy, and fluency are distinct constructs, they are also interconnected. The non-linear development of these constructs bears witness to this interconnectedness. How language is internalized and then modified (accuracy and complexity) competes with how it is accessed (fluency). This is so because there is a difference between acquiring the language and using the language as argued by Housen and Kuiken (2009) and Housen *et al*. (2012). Acquiring the language needs conscious attention to the input to internalize the rules of the language and in some cases even monitoring the produced language. However, using the language needs automatic access to the knowledge already internalized. So, these processes are in competition. That is why, as mentioned earlier, there are two opposing views about this.

6. The operationalization and measurement of CAF

Lambert and Kormos (2014:608) have emphasized that only using general measures of subordination to operationalize complexity might obscure the following important information: "(1) types of subordination, (2) item-based use of subordinate structures, and (3) potential interactions between subordination, discourse genre, and mode of production". They have given examples of these cases. As an example of the first case, considering subordination as one construct and measuring it all together will not reveal information about (1) nominal clauses as objects of superordinate verbs; (2) adverbial clauses to modify superordinate verbs; and (3) relativizing clauses to modify superordinate nouns although these might develop at different age ranges and under different conditions. An argument to exemplify this case is that it has been confirmed in research (Halliday and Mattiessen 1999; Halliday and Martin 1993 as cited in Lambert and Kormos 2014) that different types of processes appear at different proficiency levels (e.g. coordination of simple sentences with beginners, verbal subordination with intermediates and nominalization with advanced level). However, measuring subordination as only one construct may not reveal this important information. As an example of the second case, Lambert and Kormos (2014:608) claim that the clauses that occupy the position of object of verbs like *think* and *see* are not cases of subordination because "the process cannot be extended to other contexts". However, as they state, measuring them within subordination and not alone may mask such information as these clauses are very common in the production of beginners and intermediates. However,

Foster and Skehan (1996:304 see also Skehan, 2003) have argued for the use of general measures and have concluded that it is more convenient to use general measures if the "tasks do not provoke hypotheses about the use of specific forms".

As for accuracy, Lambert and Kormos (2014) have referred to errors as its basic measure arguing that they might not be a valid indicator either for L2 development or distinguishing levels of proficiency. This is because a text that is characterized by the existence of items that appear at late developmental stages in the interlanguage may be full of errors and conversely a text that is characterized by the existence of items that appear early at developmental stages of interlanguage might be quite accurate. They also referred to the fact that making "dichotomous distinctions" between correctness and incorrectness may prevent important information from emerging from the investigation made by the researcher. For example, grammaticality and acceptability are two different constructs but it is difficult for raters to take them into consideration separately while coding the data.

Norris and Ortega (2009:556) have called for "organic" and "sustainable" measures of CAF, measures that operationalize CAF as dynamic, interrelated, and multidimensional, and measures that "inform, rather than confuse, SLA research" (see Yang and Sun, 2015:306 for a practical view on this topic). They have argued that most of the measures employed for subordination as a dimension of syntactic complexity do the same job and hence they are "redundant" as the ratio of dependent clauses to a unit of production. Moreover, they highlight that measures that index length of units may capture different forms of complexity. For example, the mean length of a sentence, T-unit, C-unit, AS-unit[18], or utterance can increase in different ways, e.g. by the addition of different elements which might be subordinate clauses but they might be pre- and post-nominal modifiers in noun phrases or they might be non-finite clauses. So, these measures are only measures of overall syntactic complexity. On the other hand, the clause length measure used without specifying the kinds of clauses does not measure the amount of subordination but length of noun phrases through pre- and post-modification, or reduced relative clauses. That is why, although clause length seems to be "superficially" like the other length measures, it is a metric for phrasal complexity. Accordingly, I believe researchers can benefit from using different specific measures that can measure exactly what they intend to measure, no matter if these measures have been already used by other researchers or not. For example, one can measure the length of noun phrases by either counting pre- or post-head modifiers. Or, one can specifically measure the use of non-

finite clauses as nominals, or only measure the number of relative clauses and so on. Moreover, I would support the idea raised by Norris and Ortega that CAF have to be measured multidimentionaly. In the 16 studies reviewed by Norris and Ortega (2009), there was little focus on phrasal complexity and only 6 studies used different measures for operationalizing one construct.

Skehan (2009) has also suggested a measure of lexis that can be added to CAF. He supports his argument with some empirical findings with the native-speaker results of lexical sophistication[19] having positive correlation with syntactic complexity and the non-native speakers having negative correlation. His reasoning is that the non-frequent words did not pose any problems for the native speakers; they (i.e. the non frequent words) may have helped the native speakers to produce more complex syntactic structures than the frequently used words. However, the non-native speakers' use of less frequent words has driven their attention away from syntax resulting in less accuracy and complexity.

2.6 Conclusion

This chapter has explained learner corpora in terms of definition, design and the principle of comparison. The chapter has also explained complexity, accuracy, and fluency (CAF) in terms of background, previous research and points of debate. Learner corpora's definition is based on the general corpora's definition and their design is also dependent on a number of criteria. The principle of comparison is followed with the contrastive interlanguage analysis (CIA).

As for the origin of CAF, it was first the distinction between fluency and accuracy that emerged and then Skehan (1996) developed the idea into a "triad" model adding complexity. The research concerning CAF can be divided into two types: the research that dealt with CAF as its main focus and the research that dealt with CAF as dependent variables to measure the effect of other factors on learners' task performance. As a result, there is a lot of debate about CAF. The points of debate include the importance of the similarity as well as the difference in results, the difference between accuracy and comprehensibility, definition and operationalization of CAF, the interrelation among the three components of CAF, and the cognitive approach to CAF.

The research surveyed in this chapter demonstrates that there has been a great deal of work carried out on CAF but not all of this work involves data could be called a corpus.

For example, some of the data includes only a small number of essays or maybe isolated sentences or collections that are not structured as corpus should be. However, the work carried out by Granger and colleagues on learner corpora has demonstrated the benefits of using a learner corpus properly designed and compiled. This work has inspired me with knowledge about how to compile my corpus and the points needed to be taken into account while collecting the data of the present study and keying them on computer. For example, all the points discussed in Granger (2008) in regard to task and learner related variables were considered during the process of corpus compilation and all the needed information about these variables were obtained.

Moreover, one other major point that this chapter has indicated is that there are two types of studies that have dealt with CAF. Those studies that have measured CAF to assess the effect of other factors such as corrective feedback, planning, task type etc. is those studies that have studied CAF to decide (1) what unit is the optimum unit for analysis for measuring complexity and accuracy, (2) what measure is the best measure for assessing development in writing, (3) whether areas should be measured jointly by combined measures or separately by item-based measures e.g. measuring phrasal complexity by taking all post-nominal modifiers together or separately, and (4) which studies have checked the reliability and validity of their methods and measures of analysis. oncerning the first group of studies, it has been found out that the complexity, accuracy and fluency are affected differently by different factors e.g. feedback vs. no feedback, online-planning vs. pre-task planning. The latter group of studies that not all researchers have reported the reliability and validity of their studies and that there is not any consensus as to what unit should be used for measuring complexity and accuracy and what measures are universal. For instance, T units have been criticized, as their use is an imposition on the "psychological reality" of sentence (Bardovi-Harlig, 1992) measuring complexity it has not been confirmed that measures based on are good measures for assessing writing development (Biber *et al.*, 2011). Two other main points of discussion are that most researchers have used dichotomous decisions of either correct or incorrect units without considering a weighted unit and complexity and accuracy are in a trade-off relationship.

All these points of discussion have inspired my research in different ways. Balancing T units against sentences, I found it more appropriate to use T units though they have been criticized because sentences are found to be inadequate units for the analysis of the data of the present study as one sentence might sometimes a whole paragraph due to the

problem of punctuation confusion and . That is why I believe that what to use as a unit of analysis depends on the type of data you are dealing with and hence there is a need to address different problematic data like the data of the present study. These data are rich in concepts but very poor in language to an extent that they cannot be dealt with unless taken in a wider context, and/or analyzed by somebody who shares the learner native language and/or taken in small units. Another point that I found it important is to take account of multiple measures instead of only trying one measure and thus I decided to use different measures for one area. For example, complexity has been measured using different measures for coordination, subordination and phrasal complexity and phrasal complexity has also been measured using multiple measures involving all types of post-nominal modifiers. Also, I tried different measures for accuracy until I devised the method of IAA. As for reliability, I came out with a different method and I called it user engagement, which both involved the other teachers in the process of data analysis and helped in checking 'discussion-based' inter-rater reliability of my analysis of data. Complexity has also been compared to accuracy for a number of students to find out if the issue of trade-off relationship is true. Before bringing this chapter to an end, it important to highlight that the next chapter will discuss the process of data collection for this thesis, the methods used in the analysis of this data, and the corpus design.

.

Notes

1. Although CAF looks to be one unit and hence singular, I am going to follow convention and treat it as plural.
2. The word 'systematic' is based on Nesselhauf's (2004) definition of learner corpus.
3. CA is a method of comparing the L1 with the L2 to identify the areas of difference, which are more likely to be the areas of difficulty for the learners. See chapter 4 section 4.3.
4. I will discuss computer error analysis in chapter four.
5. In addition to the studies described in this section see also Iwashita (2006), Ishikawa (1995), Lim (1984), Homburg (1984).
6. See the section 2.4.2 on the research makes use of CAF as dependent variables.
7. A number of researchers have provided useful summaries of research studies on CAF, in addition to providing results of their own studies (e.g. Polio and Shea, 2014; Evans *et al.*, 2014; Ortega, 2003; Wolf-Quintero *et al.*, 1998; Polio, 1997)
8. See Towell (2012) for a good description of CAF within the framework of psycholinguistics and the list of the studies that have studied CAF in this regard.
9. The studies listed in Wolf-Quintero *et al*. (1998) and Polio and Shea (2014) on accuracy are tabulated in chapter five on correctness analysis.
10. This list of criteria has been used for the division of the essays of this study as well but more criteria are added to the list as deemed suitable for the division of these specific essays.
11. The measures used by Polio were error-free T-unit/ total number of T-units (EFT/TT), error-free clauses/ total number of clauses (EFT/TC), error-free clauses/ total number of words (EFT/TW).
12. These rubrics are copied from Polio and Shea's (2014:25) as they are but put in a table.
13. The measure can be calculated as follows: Word types per square root of two times the words for measuring vocabulary complexity. This is used for measuring lexical complexity.
14. The measure of weighted clause ratio has originally been developed by Wigglesworth and Foster (2008).
15. This is also a unit for measuring CAF in spoken language.
16. This is a measure used to measure lexical variety. In this measure, a given text is divided into segments of a specified number (mostly between 40-50 words) and the type-token ratio is calculated for each segment by dividing the total number of different words in that segment by the total words in the segment. Then the MSTTR is calculated for a given text by calculating the average of TTRs for all the segments. (This measure is described in chapter nine).
17. This is also a measure for lexical complexity.
18. AS unit (analysis of speech unit) is also a unit that is used in the analysis of spoken production. It is defined and explained in chapter 5.
19. Measured with lambda, a measure for lexical sophistication.

Chapter Three

Data Collection, Corpus Design and Methods of Data Analysis

3.1 Introduction

This chapter gives an account of the process of data collection, corpus design and methods of analyzing the data used in this study. It answers questions such as how did the sampling take place? And why has such a sample of subjects been chosen? It also explains the process of ethical review approval as an important part of the data collection process. Moreover, it gives a full account of the task used to elicit the writing from the subjects and the way this writing is computerized and turned into a raw corpus. It ends with a description of the methods used for carrying out this research.

3.2 The sample

Since the main research question to be answered through this research is: how can the accuracy and complexity of *the written English* of *Kurdish university students* be measured, Kurdish university students majoring in English, who are supposed to represent advanced learners of English, were chosen as the sample of subjects and the essay writing was selected as the research task.

The sample comprised third and fourth year students at three English departments at the School of Arts, Basic Education, and Languages at two universities (University of Dohuk and University of Zakho) in Dohuk province in the Iraqi Kurdistan region. This is an autonomous region in the north of Iraq, where Kurdish (with its two dialects: Sorani and Bahdini) is the first language of the majority of its population, Arabic the second and English the dominant foreign language. Great importance is now being attached to teaching English in most of the public and private teaching institutions in the area.

The choice of these two universities was mostly opportunistic due to the fact that (1) I work at the English department of one of these universities (University of Dohuk, School of Arts) (2) I had personal contacts in these universities and so had ready access to staff and students (2) all the staff (including the teaching staff and the administration) of the two universities expressed their readiness to cooperate with me in all ways to facilitate the process of data collection, (3) the heads of the English departments were very cooperative in providing me with written approval[1] to take their students as subjects of

the study (4) most students also expressed their willingness to take part in the task I gave them, and (5) the three groups of students (at the three English departments) are homogenous, especially in terms of their mother tongue and the Kurdish dialect they speak, (Kurdish/Bahdini) and age range (19–40)[2]. The data collection process and corpus design took place during November and December 2014 and lasted for 40 days.

I selected the third and fourth year students because they were the most advanced students available and I wanted to compare the English written by the fourth year students to the written English of the third year students. This is to measure if there is any improvement from level three to level four. I chose the three schools because I wanted my research to cover a range of students not only at the institution where I work but two schools as well where the modules taught are different and the teaching style is also different. I did not want a totally homogenous corpus in terms of text type and thus I chose two types of essays (argumentative and narrative). Other aspects of selection were not under my control. For example, I could not keep a balance between males and females in each group (third or four, or even argumentative and narrative) because . Also, the students in each group (third and fourth) at the three English departments were not at homogenous proficiency levels. A number of the third year students may get 5 in the writing component of IELTS but other may not get more than 2. The reason I have not divided them into levels according to a proficiency test is that was beyond the scope of this research and my only purpose was to compare year three to year four and hence I considered year of study as an external proficiency indicator rather than having an internal linguistic measure.

These students have studied English for eight years before joining university because teaching English at the time when these students were at the primary schools used to start at grade five. However, English is taught at the primary and secondary stages as one single module and the rest of the modules are taught in Kurdish. Third year students have been studying English for another three years and fourth year students for another four years at university where they specialize in English language and literature and all modules are taught in English and about English language and literature. In this way I had control over what level of students to choose but I had only partial control over the number of the students who should participate in the research because the availability of students at the time the task was conducted and the ethical issue (students' consent to take part in the task) restricted the number (See table 3.1 for the number of the students that

participated in the task) and did not give me the opportunity to have equal number of students in each group though that was intended.

Table 3.1 a summary of the sample of the present study

School of Arts						School of Basic Education						School of Languages						T
Third year			Fourth year			Third year			Fourth year			Third year			Fourth year			
N	A	T	N	A	T	N	A	T	N	A	T	N	A	T	N	A	T	
31	26	57	27	28	55	30	26	56	26	26	52	21	20	41	24	23	47	308

N= Narrative essay A=Argumentative essay T=Total

As table 3.1 shows, from the School of Arts (University of Dohuk), 57 (24 male and 33 female) third year students and 55 (21 male and 34 female) fourth year students participated in the study. As for the School of Basic Education (University of Dohuk), 56 (22 male and 34 female) third year students and 52 (25 male and 27 female) fourth year students performed the task. From the School of Languages (University of Zakho) 41 (17 male and 24 female) third year students and 47 (16 male and 31 female) fourth year students took part in the study. Only one of the essays was excluded as the student had solely copied the topics repeatedly without writing anything else, even though he/she agreed to participate in the study and signed the consent letter.

Table 3.2 the background of the subjects in terms of the L1 and foreign languages other than English

Levels	Mother Tongue				Other Foreign Languages Besides English				
	Kurdish/ Bahdini	Kurdish/ Sorani	Arabic	Syriac	Arabic	Arabic and Other Foreign Languages	Other Foreign Languages Only	No Foreign Language (only English)	Lived in English Speaking Country or EU countries
3rd year Arts	53	2	1	1	26	1(Persian Turkish), 2 (French), 1 (Turkish) 3 (Persian), 1 (Dutch, Swedish), 1 (Indian)	1 (Turkish), 1 (Korean), 1(Norwegian), 2 (French) 1 (German)	16	1 in Spain for one year and six weeks in the USA, 1 was born in the Netherlands and lived there for a long time
4th Arts	49	1	2	3	27	5 (French) 1 (French, Turkish and Indian) 1 (Persian, Turkish and	1 (French), 1 (Turkish) 2 (Kurdish as Arabic is the L1 language)	16	
3rd year Basic	50	6	0	0	41	1 (Persian), 1(Turkish)	2 (Persian), 1(Turkish), 1 (Norwegian), 1 (Indian)	9	
4th year Basic	42	5	1	4	23	2 (Indian), 2 (Kurdish as Syriac is their first language), 1 (Turkish –Persian), 1 (Persian), 1 (Turkish-French)	1 (Persian)	21	
3rd year Languages	35	2	3	1	25	1 (Kurdish as Syriac is his mother tongue)	1 (German, French, Chinese), 6 (Turkish), 1 (Persian)	7	one was grown up in Germany
4th year Languages	44	2	1	0	28	4 (Turkish), 1 Persian, 1 (Indian)	1 (Persian-Turkish), 4 (Turkish), 1 (Kurdish as his mother tongue is Arabic	7	

As per table 3.2 except for 9 students who spoke Aramaic (Syriac)[3], 8 who spoke Arabic, and 18 who spoke Kurdish/Sorani as their first languages, all the remaining students spoke Kurdish/Bahdini as their first language. Besides English, other foreign languages spoken by the subjects included Arabic, Turkish, Persian, Indian, Korean etc. Very few who had lived in EU countries like Germany, the Netherlands etc. spoke Dutch, Swedish, French, German and Chinese as foreign or second languages.

Additionally, very few of the students had lived in English speaking countries and the EU countries. One female student at the University of Zakho has grown up in Germany. Another female student at the University of Dohuk had spent only six weeks in the USA, whereas another female student has grown up in the Netherlands.

3.3 The tasks

Because I did not want a corpus that is too homogeneous i.e. I wanted more that one essay type but at the same time I did not want a corpus that is completely random in terms of register and genre as well, I specified two essay types (argumentative and narrative). I gave two tasks to the students: writing argumentative essays and narrative

essays to get an equal number of argumentative and narrative essays. However, as stated earlier it was under my control to choose the task type and the text type but it was not under my control to balance the number of students in each group. The rationale behind why I choose these two genres is that I expected that the argumentative essays would be more difficult for students, they would be less familiar with this genre and they would require a greater command of complex sentences. This is because they are mostly 'an agree and disagree' position which needs evidence and counter-evidence for supporting an argumentative status. The narrative essays on the other hand, I expected, would be less complex and would be with a high number of verbs, as they will be consisting of a series of events. In other words, I expected that the two tasks would contrast in many ways. These tasks were controlled for time, topic, and length. They were two-hour essay-writing classroom activities with the length of 400–500 words each. The following table shows the ten topics that were given to the students.

Table 3.3 the ten topics that were given to the students.

	List A: Argumentative essays	List B: Narrative essays
1	Do you agree /disagree that the countryside life is much healthier than the city life?	Your first week at college
2	Do violent video games cause behavior problems?	The funniest or saddest event in your life
3	Should cigarette smoking be banned?	The day you met your best friend
4	Do you agree/disagree with the suggestion that English become the language of education in Iraqi Kurdistan region? All subjects be taught in English.	Any of your childhood experience that you will never forget
5	The advantages and disadvantages of single-sex schooling and co-educational schooling.	An unforgettable journey

The list of topics is divided into two: one list of narrative and the other one of argumentative topics. This is because the students at each level (year) of study were divided into almost two equal groups (although that was hard to manage, see table 3.1) and one group was given the list of narrative topics and the other the list of argumentative topics to choose one topic from and write an essay about. Thus, the tasks were partially controlled for topic, as I wanted to give the student some choice in selecting what to write about. More importantly, the topics were chosen with care. They are everyday topics and

directly related to students' lives. No political, religious, and sensitive social topics are included.

The task was not given to the students in the form of an exam (i.e. to be later corrected and marked) but rather it took place as a normal classroom activity mostly in the writing classes as required by the heads of the selected English departments. However, the students were not given the opportunity to use any reference tools like dictionaries or any search engines such as Google. Also, the task was administered with invigilation. The students were not allowed to copy each other's ideas or words because the original aim of the study is to compare individual essays in terms of accuracy and complexity.

Attention was paid to ensuring students understood the assignment and that the groups were treated equally. For example, the topics were read to the students and explained to them both in English and Kurdish. The students were allowed to ask questions for further explanation of the topic but no assistance was given to them with the content of their essays i.e. they were not helped with vocabulary or sentence structure. Although the other teachers gave me help in invigilating the task, I took full responsibility for giving the necessary instructions to the students, and only I read and explained the topics to them. It is also worth noting that where possible students of a similar level (third or fourth year) performed the task on the same day; if this was not possible the list of topics was removed at the end of the task. This was to avoid students showing the topics to others and to avoid students bringing 'ready-made' essays on the day the task was carried out. In order to keep students' personal information confidential, the students were informed not to write any of this information on the essay but instead were given separate sheets of paper for that purpose.

3.4 Ethical review

In adherence to the University of Birmingham's Code of Practice for research, all research studies involving human subjects[4] are subject to an Ethical Review Approval from a specialized Ethics Committee. As stated in this Code of Practice, an ethical review form designed for that purpose must be filled in and sent to the Ethics Committee for approval. This form[5] requires detailed information such as the title of the study to be carried out, a summary of the project to be undertaken, the methodology the researcher is going to use for data collection and analysis, and a breakdown of the number, age, gender, location, intellectual ability etc. of the subjects to be involved. Ensuring

confidentiality or anonymity of the data, participants consent, the required action in case of participants' withdrawal, and the feedback given to the participants upon completion of the study are all included in the form. This is in addition to the research significance and benefits.

This process of obtaining the ethical approval was completed for this study and a copy of the completed form is attached as appendix D. As per this ethical review approval, the subjects were informed both verbally and in writing that all their personal information would be kept confidential. Confidentiality was maintained by giving the students a separate sheet of paper to record their personal information. Later, these sheets and the essay papers were given the same codes and then the sheets of the personal information were detached from the essay papers and kept for further reference while designing the corpus.

Every student who agreed to participate in the research project was required to sign a letter of consent[6] detailing all the information needed to be known by the students such as the task type, the length of the essay, the time allowed, the issue of confidentiality, and notification of the results to the students upon completion of the study. Moreover, the letter contained my contact information and that of my supervisor in case the students wanted further clarification. As I thought it did not suffice only to give the letter to the students without reading and explaining it to them, the letter was explained to them both in English and Kurdish and if needed, in other languages, such as Arabic.

The process of obtaining the ethical review approval included obtaining written consent, as mentioned earlier in section 3.2, from the heads of departments whose students participated in this research project. The letter showed the heads of the departments' full approval of their students taking part as subjects in the study.

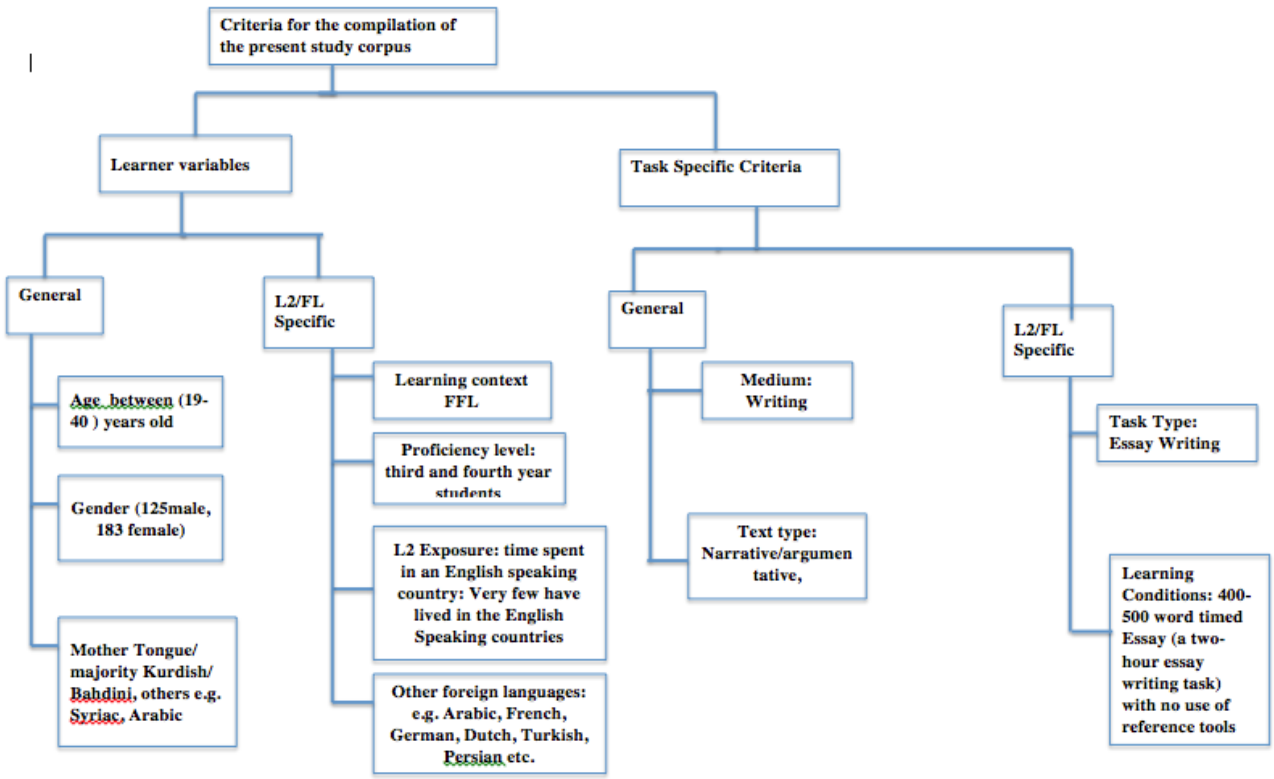
3.5 Corpus design

It is important to highlight that only a number of Granger's (2008) criteria in figure 2.1 chapter two have been taken as standard points or decided before data collection. For example, I decided in advance that I would only take third and fourth year students at the Universities of Dohuk and Zakho. Also, I decided that the activity to be given to the students should be a timed in-class essay on one of five narrative or five argumentative topics. Moreover, I decided that the students would not be supported in terms of language

support materials such as a dictionary. However, some other points in Granger's (2008) criteria such as age, gender, knowledge of other foreign languages besides English or time spent in an English speaking country have only been taken as important information to collect along with collecting the data.

Taking this into consideration and although the selection of participating institutions and participants was opportunistic (see section 3.2), still the corpus compiled can be described in terms of figure 2.1 chapter two. Other than the level (year) of study, the points taken into account[7] are (1) learner/student specific variables that are further divided into (a) general variables such as age, gender, and mother tongue of the student and (b) second/foreign (S/F) language specific variables such as the learning context (whether it is English as a foreign language (EFL) or English as a second language (ESL), proficiency level (considered here as year of study), knowledge of other foreign languages in addition to English, and time spent in an English speaking country, and (2) task specific variables, which are also further divided into (a) general variables such as medium and text type and (b) S/F language specific variables such as task type and learning conditions. The following figure shows all the points taken into account while compiling the corpus of the present study.

Figure 3.1 information about the corpus of the present study



The essays were keyed into computer by a group of colleagues who were asked to retain all the errors originally made by the subjects. I later checked the keyed essays against the original ones to ensure that accuracy was kept to the maximum while keying the essays.

If one compares this process of corpus compilation to the definition of learner corpus in chapter two, would the resulted corpus qualify as a corpus of learner language? The following is a point-by-point check:

1. Electronic collection: the essays are electronic as they are keyed into computer.
2. Authentic: authenticity in FL/SL is considered as a continuum and "in as far as essay writing is an authentic classroom activity, learner corpora of essay writing can be considered to be authentic written data" (Granger, 2002:8 see section 2.2 on the definition and design of learner corpora).
3. FL/SL textual data: the essays are continuous stretches of language produced by Kurdish learners of EFL as an in-class activity, meaning they do not contain sentences out of context.
4. Explicit design criteria: the learner-specific variables (age, gender, knowledge of foreign languages other than English, time spent in an English speaking country etc.) and task-specific variables (time, topic, use/no-use of reference tools) are taken into account.
5. SLA/FLT purpose: the purpose of compiling these essays is to measure the accuracy and complexity of English writing of Kurdish students. This is a clear SLA and FLT purpose.
6. Standardized and documented in a homogenous way: the texts are computerized as plain texts (raw corpus). They are marked for school and level (year) of study and tagged for errors and correct forms.

Other issues that remain to be further taken into consideration while designing a corpus are the representativeness, balance and size (Hunston, 2008). As for the design of the present corpus, serious attempts have been made to include all the third and fourth year students of the three English schools at two universities in the research but the availability of the students at the time of the task, the original number of students in each class, and the ethical matters all played a major role in the size of the corpus and the balance between the small components of the corpora (the small-sized sub corpora). As for representativeness, the corpus only represents the language produced by the third and fourth year students majoring in English in three different schools at two universities in

Iraqi Kurdistan. This is why this corpus can only be "judged in this context" (Hunston, 2008, see the full quotation in section 2.2).

All the conditions that characterize any data as learner corpus are present in the data of this study and hence this data can be considered as a specialized learner corpus compiled for an SLA/FLT purpose, namely measuring accuracy and complexity in the written English of the Kurdish students learning English in an EFL context.

3.6 A summary of methods of data analysis

For achieving the aims of the present study, the following procedures have been followed:

1. Different methods have been used to measure accuracy, and grammatical and lexical complexity, and analyze the data.
2. When the measures did not work well, an attempt has been made to explore new metrics of accuracy and complexity.
3. An attempt has been made to use automated ways of measuring accuracy and complexity through using corpus tools.
4. For each chapter of data analysis, 20% of the data has been reanalyzed by a second rater in a discussion with me as the first rater. The second rater (or better to call him/her *a user)* is specialized in Applied Linguistics and is very likely to use the research. A different rater has worked on one chapter at a different time. The discussion sometimes took one hour for each essay. Although this process was not easy, it saved both time and effort. The complexity of the data and the various possible ways of analysis would have needed a lot of time and effort by many people if they were reanalyzed independently. Therefore, having a second rater or user reanalyze the data in discussion with me has been both useful and less time-consuming. One of the outstanding benefits of this method is that when possible inaccuracies are based on arguable issues like collocation and lexis, discussion is more useful that independent coding. Moreover, one can summarize the aims of this method in three points: (1) checking the accuracy of analysis in a discussion rather than independently, (2) training teachers in the methods of analysis, (3) engaging their interest in using these methods, especially the new ones. The method is hence called 'user engagement'.

Notes

1. The approval letters of the three departments are attached as appendices A (approval letter from the Department of English, School of Arts, Faculty of Humanities, University of Dohuk), B (approval letter from the Department of English, School of Basic Education, Faculty of Education, University of Dohuk), and C (approval letter from the Department of English, School of Languages, Faculty of Humanities, University of Zakho).
2. Although the actual age range was 19–40, the age of great majority of the students ranged between 19–22 i.e. only one student was found to be in her early forties and two in their early thirties.
3. Visit http://www.nineveh.com/whoarewe.htm for a description of the Assyrians who speak syriac as their first language.
4. And even animals to ensure that the benefits of research are more than and/or worth the effects that they will undergo (see https://intranet.birmingham.ac.uk/as/registry/legislation/codesofpractice/index.aspx for more information)
5. available on https://intranet.birmingham.ac.uk/finance/accounting/Research-Support-Group/Research-Ethics/Ethical-Review-Forms.aspx
6. Attached as appendix E.
7. A copy of the sheet of criteria is attached as appendix F, and as mentioned earlier, these criteria were all recorded on a separate sheet of paper for matters of confidentiality.

Chapter Four

Error Analysis

4.1 Introduction

This chapter covers one important method of measuring accuracy in language, namely Error Analysis (EA). It first describes EA in terms of theory and application. It also presents the points of criticism that have been raised by different scholars against EA; these criticisms have not rendered EA obsolete but have changed it from its traditional form to its computer aided form i.e. computer-aided error analysis (CEA). This chapter therefore describes CEA as well giving Université Catholique de Louvain's error taxonomy and error tagger as examples. It also describes the Dexter Coder as a computer tool used for error tagging the data of the present study. It then presents and discusses the results of applying CEA to the data of this study. It concludes with an argument that EA on its own cannot function as a good measure of accuracy, especially with low-quality writing.

4.2 EA as a measure of accuracy

The concept of error and error-free units has dominated the research that has aimed to measure grammatical accuracy as one of the three dimensions of language proficiency and development (i.e. fluency, accuracy, complexity, see Housen and Kuiken, 2009; Skehan, 2009; Larsen-Freeman, 2009 for a detailed explanation of the three dimensions and interrelation between them and also see chapter two)[1]. Grammatical accuracy or "the ability to be free from errors while using language" (Wolfe-Quintero *et al*. 1998: 33), like the other two dimensions, has first been used as a "dependent variable" (Housen and Kuiken, 2009:462) with the aim of determining the role of other factors in the written and spoken learner language (see chapter 2 section 2.4.2 for a detailed discussion on this topic). Also, as mentioned earlier grammatical accuracy was also measured for its own sake rather than for the sake of measuring the effect of other factors. Polio (1997), Polio and Shea (2014) and Evans *et al*. (2014) are good examples of research that involves such measurement.

To identify improvement in grammatical accuracy, the previous studies on accuracy used different measures, all involving error in one way or another. The measures included

holistic (e.g. Abdollahifam, 2014; Marzban and Arabahmadi, 2013) and analytic measures. The analytic measures comprised the ratio of the total number of some specific correctly supplied grammatical structures or items to the total number of obligatory occasions of these structures and items expressed by the proportion of 1 (e.g. Shintani *et al.* 2013), total number of errors per 300 words (e.g. Riazantseva, 2012), total number of errors per 100 words (e.g. Chandler, 2003), total number of whole errors to the number of words (e.g. Polio, 1997), the number of error-free units[2], such as T-units and clauses to the total number of clauses or T-units or the ratio of weighted error-free T-units or clauses/ total number of T-units or clauses (e.g. Polio and Shea, 2014; Evans *et al.* 2014; Rahimpour and Hosseini, 2010; Polio, 1997; Casanave, 1994). As will be evident, the use of any analytical method involves breaking down the text into small units such as clauses, T-units and C-units etc. and these units will be discussed in chapter five.

4.3 Historical and theoretical background of EA

Although the concept of erroneous language as compared to the norm, i.e. the native speaker language, dates back to as far as the 18th century with the prescriptive grammarians (R. Ellis and Barkhuizen, 2005), EA as a methodological framework for analyzing learner language was introduced only in the late sixties, by S. Pit Corder with first his influential article 'The Significance of Learner's Errors' in 1967 and later with his other works mostly written in the early seventies (Troike, 2012; R. Ellis and Barkhuizen, 2005; Jordan, 2004). Corder's (1967) point of view about learner's errors was indeed a turning point in the role of error in language learning. Instead of viewing errors as sins that must be avoided by the language learner, Corder (1967) attached great importance to them, emphasizing that they provide insights into the nature of learner language i.e. interlanguage (a term coined by Selinker in 1972) and the mental processes underlying it. More importantly, for Corder (1967:167) learner's errors are significant for three reasons: first, in relation to teaching, they are good indicators of a learner's development in their route towards the target language; second, in relation to learning, they are used by the learner for "testing his [sic] hypotheses about the nature of the language he is learning" and third, in relation to research, by investigating them the researcher gets a picture of "how language is learned" through identifying the properties of the language system the learner is using (their 'transitional competence'[3]), which is unique in the sense that it is neither similar to their first language nor the target language they are attempting to learn.

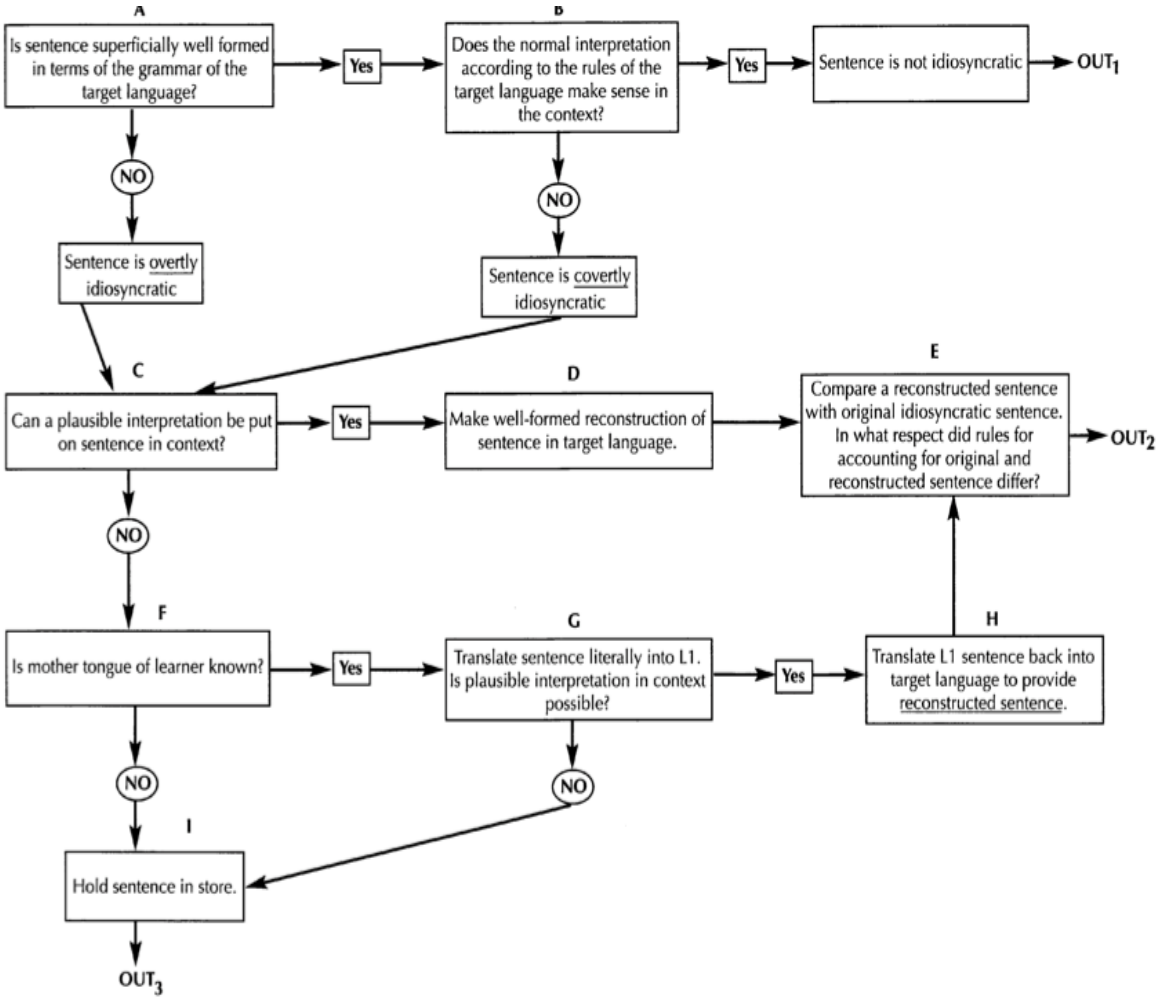
The EA emerged and flourished as a counter theory to the transfer theory of Contrastive Analysis (CA). CA's main concern was the comparison of the 'habits' of the L1 with those of the L2 (see N. Ellis, 2008; Gass and Selinker, 2008; R. Ellis and Barkhuizen, 2005; de Bot *et al*. 2005; R. Ellis, 1994). Moreover, its main role was mostly pedagogical as it aimed to compare the native language of the student (i.e. learner) to the target language, predicting the areas of difficulty that students face while learning the target language so that these are given more attention in teaching. CA equated difficulty in language learning to the difference between L1 and L2 giving the indication that all differences will result in hindering ('negative transfer') the process of learning in which the learner is engaged and hence in errors (Gass and Selinker, 2008; Trawinske, 2005; see Wardhaugh, 1970 for a comprehensive review of CA). Although CA and EA both focused on errors, the way they viewed and studied them was different. CA viewed errors as "sins that needed to be avoided at all cost" (Johnson, 2004:25), while EA considered them as "windows" into the learner's mental language system, through which the researcher can describe the rules underlying learner's language (see Troike, 2012; S. Johansson, 2009; Gass and Selinker, 2008; Corder, 1967). It is worth mentioning that CA exists in two versions: the strong version and the weak version (see Gass and Selinker, 2001). As pointed out by Gass and Selinker (2008) and Johnson (2004), CA compared only L1 to L2 in its strong version and learner language to L1 in its weak version but EA compared only learner language to L2.

4.3.1 Application of EA: identification and description of errors

4.3.1.1 Identification of errors

To identify errors in a sample of learner language, Corder (1981:23) [4,5] provides the following model:

Figure 4.1 Corder's (1981) model of error analysis



In figure 4.1 Corder presents an algorithmic model for identifying errors. According to Corder (1981:21) "every sentence is to be regarded as idiosyncratic unless shown to be otherwise", meaning that every sentence is wrong unless proved to be right. He distinguishes between covertly erroneous utterances (well-formed at the sentence level but not explicable in a wider context) and overtly erroneous sentences (clearly ill-formed at the sentence level). Therefore, he suggests a reconstruction of the sample in terms of native speaker language as a norm. What he actually meant by 'reconstruction' is the reproduction of the sentences in the native speaker language (L2), if they can be 'interpreted'. If they cannot be interpreted, Corder suggests resorting to the mother tongue of the learner for interpretation, meaning that the sentence be reconstructed in learner's mother tongue and then translated back into the native speaker language. The reconstructed version can later be compared with the original sample for identifying errors.

This model looks to be a very well organized framework for conducting EA. Its

application, however, is not without problems. First, as questioned by R. Ellis (1994), which variety of the target language is meant? Written or spoken? American, British, Zambian or what? Errors in speech, for example, cannot be identified as deviations from the norms of the standard written variety. Also, Ellis argues that (1994:51), a sentence like *she coped up with her problem very well* will obviously be incorrect if the norm of comparison is the standard written British or American English but it is correct if the norm is educated Zambian English. Nevertheless, an error analyst can define what an error is in their context and can provide the norm as well. Or, the definition provided by Lennon (1991:182) "a linguistic form or combination of forms which, in the same context and under similar conditions of production, would, in all likelihood, not be produced by the speakers' native speaker counterparts" might work well in all situations because of two key phrases, namely *same context* and *under similar conditions of production*.

As highlighted in the literature (see Brown, 2007 and R. Ellis and Barkhuizen, 2005), even if an unambiguous definition is provided, there remains the problem of error and mistake. Corder (1967, 1981) makes a distinction between mistakes and errors emphasizing that the former are not due to the learner's lack of knowledge of the native speaker language because they might even be made by native speakers themselves who have full knowledge of the language but the latter are the result of learner's lack of knowledge of the native speaker language. In the application of EA, it is, however, not straightforward to distinguish between the two. This is because Corder seems to be doing two distinct things here. One is labeling the text and the other is interpreting that labeling in terms of cognitive development. In other words, he mixes the description of errors and identification of their cause. Probably doing the two things at once is what causes the trouble. Assuming that such a process of differentiation between errors and mistakes is valid and possible, scholars like R. Ellis and Barkhuizen (2005) and Brown (2007) have suggested the frequency of the erroneous form as a criterion. For example, if an erroneous form appears consistently throughout the whole sample of a given learner language, it can be considered an error. To give Brown's (2007) example, if a learner has alternated between forms like *John can sing* and *John cans sing,* it is suggested that the analyst makes a further analysis of the sample of that learner's language. If they have produced more tokens of the same type such as *John wills sing*, *John mays sing*, it is then easy to decide that *John cans sing* is an error not a mistake. But, what if some errors/mistakes like lexical errors appear only once or no more similar errors occur throughout out the whole text? The researcher in this case will be confused whether it is an error or a mistake and has to decide based on their personal intuition.

Other problems might also arise from the actual application of the model, which complicates the identification of errors. For example, even if the answer is YES to question C in the algorithmic model in figure 4.1 "can a plausible interpretation be put on sentence in context", the researcher comes across ambiguous cases. The example below (4.1) shows one of these cases:

e.g. 4.1

He did not know the word so he asked a dictionary (Corder, 1981:23) (See also the example provided by R. Ellis and Barkhuizen, 2005:59)

Corder (1981:23) mentions two possible interpretations for this sentence

a. *He did not know the word so he asked for a dictionary.*
b. *He did not know the word so he consulted a dictionary.*

Or even sometimes the error looks superficially easy to identify but a closer look at the sentence might lead to a different decision. Schachter and Celce-Murcia (1977:445) gave an example of an error that looked to be a missing relative pronoun though it was more likely due to the attempt of the learner to introduce a topic followed by a comment.

In all these cases, the solution might be, as suggested by Corder (1981:22), the "authoritative interpretation" i.e. to ask the learner what they actually intended to say but this, as R. Ellis and Barkhuizen (2005:59) describe it, is "impractical" as the learner might not be always available and even if they are, the reconstruction they provide for their sentence might not always be the one they actually meant. This is because, to quote R. Ellis and Barkhuizen (2005:59) based on James's (1998) viewpoint "errors are often indeterminate".

In figure 4.1, the problem becomes even more complex, when the answer to question C is NO because it will lead to question F, and whatever the answer to question F might be, it will not be without problems. For instance, if the answer is NO, the sentence will remain unanalyzed and even if it is YES, the resort to the learner's mother tongue is probably not reliable. This is because as Corder (1981) argued, when a researcher interprets an erroneous sentence of the learner using the learner's mother tongue when they are not there, the interpreted version of the sentence will only depend on researcher's intuition and the erroneous nature of the learner language might not be "solely explicable in terms of his mother tongue; it may be related to how and what he was taught" (1981:24).

Lastly, the problem becomes insoluble when the answer to question G in the figure 4.1 is

NO. Consider the following example analyzed by Brown (2007:262) using Corder's algorithm in figure 4.1:

e.g. 4.2 "*The different city is another one in the another two*"

Question A. Is this sentence superficially well-formed in terms of the grammar of the target language?

Answer: NO

Question C: Can a plausible interpretation be put on this sentence in context?

Answer: NO

Question F. Is mother tongue of the learner known?

Answer: Yes, Spanish.

Question G. After translating literally this sentence to Spanish, is plausible interpretation in context possible?

Answer: No plausible translation or interpretation

I. In this case the sentence should be left unanalyzed or OUT_3.

The negative answer to question G above might be due to the fact that errors are analyzable within the two dimensions suggested by Lennon (1991:191): the "domain" and the "extent". The domain is the "the rank of the linguistic unit which must be taken as context in order for the error to become apparent" and the extent is "the amount of linguistic context which the speaker needs to refashion in order to repair the error" (Lennon, ibid). As noted by R. Ellis and Barkhuizen (2005), the wider the domain and/or extent are, the more difficult the sentence becomes to reconstruct and the error to be subsequently identified. It is also worth noting that the wider the extent and/or the domain of an error are, the more "global" it (i.e. the error) might become and vice versa; the narrower the extent and/or domain of an error are, the more "local" it might become. Burt and Kiparsky (1974:73) define global errors as those that "violate rules involving the overall structure of a sentence, the relation among constituent clauses, or, in a simple sentence, the relations among major constituents" and local errors as those that "cause trouble in a particular constituent, or in a clause of a complex sentence". However, Brown (2007), besides describing global and local errors in terms of structural deviations,

relates them to the degree of comprehensibility by the reader/hearer of the whole or part of the message, the relation which Corder (1975:207) argues is "far from being understood".

4.3.1.2 Description of errors

Assuming it is possible unambiguously to identify errors, the next step is to describe and classify them. James (1998:95) argues that an error classification system should have two characteristics. It has to be (1) "well developed and highly elaborated", (2) "self-explanatory and easily learnable" i.e. "user-friendly". Two descriptive error taxonomies have been reported in the literature: the linguistic taxonomy and surface structure taxonomy (see R. Ellis and Barkhuizen, 2005; James, 1998; Johnson and Johnson, 1998; R. Ellis, 1994). The linguistic taxonomy, as explained by R. Ellis and Barkhuizen (2005), James (1998), Johnson and Johnson (1998) and R. Ellis (1994), is used to describe which level (phonology, morphology, grammar, lexis, text, discourse) of the language is affected by the error. It then extends further to which class of that level is affected. For example, if the error is in grammar, the classes affected might be the verb, noun, preposition, adjective etc. In other words, the taxonomy deals with the basic sentence structure. For more delicacy, it can further be subdivided to include categories describing which aspect of that class is affected; for instance, if the error is a verb error, the problem might be in tense, aspect, concord or auxiliary, regular/irregular verb etc. Thus, an error might be classified as Verb-Past Simple Tense-Regular/Irregular.

Unlike the linguistic taxonomy that deals with *which part* of the sentence is affected, the surface structure taxonomy, deals with *how this part* of the sentence is affected. R. Ellis and Barkhuizen (2005) list with examples the categories that this taxonomy includes. They note that when it was first proposed by Dulay, Burt and Krashen (1982), it consisted of four categories but James (1998) added a new one. The categories are (1) omission (omission of an item when its use is obligatory e.g. omission of the indefinite article in *he is very nice man*), (2) addition (the addition of an item that would not be used in the target language) which include regularization (e.g. teached for taught), double-marking (e.g. he didn't sent) and simple additions (i.e. additions other than regularization and double-marking), (3) misinformation, (the use of the incorrect form of a morpheme or a structure) which include regularization (e.g. do they be happy?), archi-forms (e.g. the use of *that* in all cases where any of the four demonstratives this/that, these/those is needed) and alternation (e.g. the use of don't+verb and no+verb), (4) misordering (the use of the

wrong word order (e.g. he told me where was he), and (5) blends, the category proposed by James (1998) (when two forms exist and the learner is uncertain which one to use, so they mix the two; James gives the example *according to Erica's opinion*, which results from the two forms *according to Erica* and *in Erica's opinion*).

It is worth noting that the two taxonomies can be used together. R. Ellis and Barkhuizen (2005:63) give many examples of such a case: *the big of them* can be classified as error in noun phrase/adjective/comparative form according to the linguistic taxonomy and error of misinformation-regularization according to the surface structure taxonomy.

4.3.1.3 Other criticisms of EA

Besides the above-mentioned problems, which accompany the actual process of applying EA, other points of criticism have also been raised by different scholars. The first one concerns the learner language data. Dagneaux *et al.* (1998:164) argue that, basing their argument on R. Ellis' (1994) viewpoint, in many EA studies the samples are "heterogeneous" as they are not collected according to "well-defined" criteria making it difficult to identify what types of learners made what kinds of errors and under which conditions. Another drawback of EA (see Granger, 2009, 2008, 2002; Dagneaux *et al.* 1998; Schachter and Celce-Murcia, 1977) is that it only focuses on errors, ignoring the other component of learner language, namely the correct forms. In other words, the errors were decontextualized. It is not only the correct forms that EA does not look at but also the forms that the learner has probably avoided, under-used and/or over-used. For example, Schachter (1974)[6] has found that Persian and Arab learners of English make more errors in relative clauses than Japanese and Chinese learners. However, Schachter has also noticed that the number of instances of relative clauses used by the Persian and Arab learners were actually much higher than those used by the Japanese and Chinese and hence more errors in the Persian and Arab learners' English. Schachter, thus, argued that if EA was only carried out on the Japanese' and Chinese' use of the English relative clauses, the analyst would conclude that they had no difficulty producing the structure. Also, in the same study, by only looking at these errors out of context, one could argue that the Japanese and the Chinese have better acquired the form. In this case, EA would have failed to capture the problem of avoidance because it did not consider all the instances (both correct and erroneous) of relative clause use in the sample.

Although EA has been recognized as an approach with all these limitations, "recognizing

[its limitations] does not necessarily spell its death" (Dagneaux *et al.*, 1998:163). On the contrary, it led to the introduction of a new EA, this time supported by the power of computer: Computer-Aided Error Analysis (CEA).

4.4 CEA

CEA resembles, as highlighted by Granger (2009, 2004), the traditional EA in that it also involves the detection, correction and analysis of learner errors but it is unlike it in that (1) it does not concentrate only on errors but also correct forms like the case with 'obligatory occasion analysis' proposed by Brown (1973) where both learner's correct and incorrect forms are analyzed, and (2) it studies error in context.

In CEA the computer learner corpus under study is first tagged for errors and their possible corrections through using an error-editor (Granger, 2008). Then, these error tags can be retrieved using a computer retrieval tool and the error types can be counted and seen in context. An example of CEA is the one developed at the Université catholique de Louvain (well described in Dagneaux *et al.*, 1998) where the learner corpus is first manually corrected by a native speaker who also inserts the corrections in the text. Then, the analyst, who is preferably a non-native speaker with a very good knowledge of English and shares with the learner their mother tongue inserts the error tags, which they choose from a detailed manual of error tags (see section 4.4.1.1 for the description of the error tagging taxonomy). The error tags are inserted using an "error editor", which is software specially made for the purpose to speed up the tagging process (see figure 4.2 below). The errors can then be retrieved, seen in context, counted etc. using computer retrieval software (see figure 4.3 showing an example of a search for the error tags verb dependent preposition and count/non-count nouns).

Figure 4.2 a screenshot as an illustration of the use of error-editor (Granger, 2002:20)

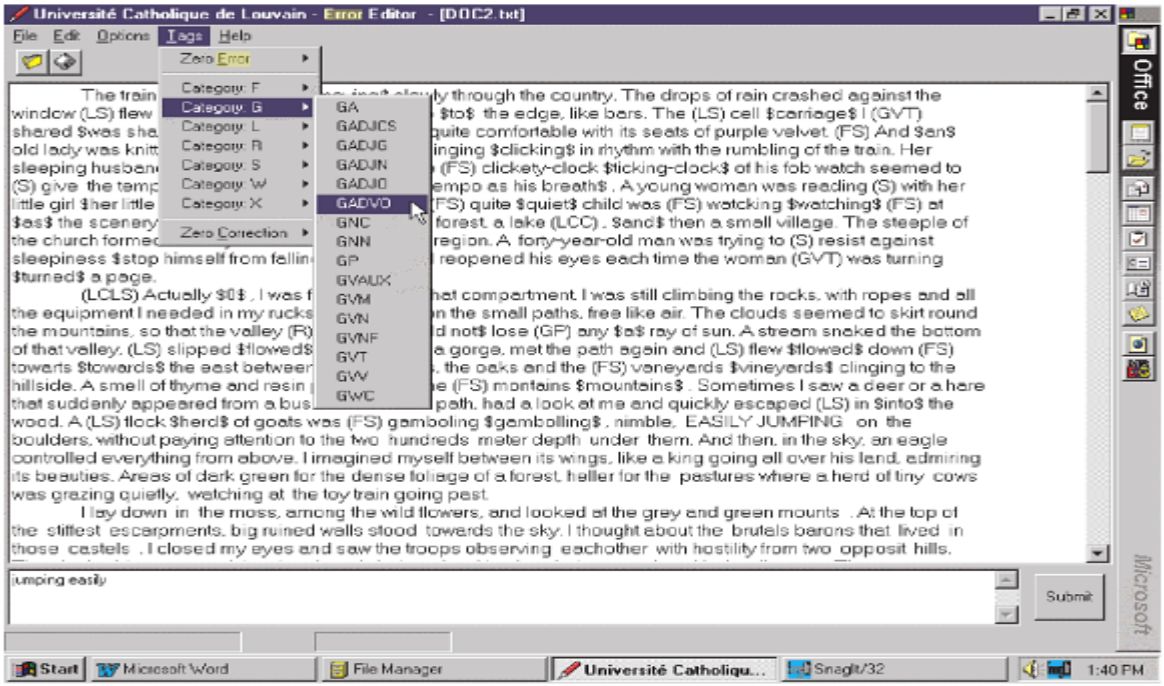


Figure 4.3 an example of the retrieval of the error tags verb dependent preposition and count/non-count nouns (Granger, 2002:20)

the fact that we could	(XVPR) argue on $argue about$ the definition of
want to be parents, do not	(XVPR) care of $care about$ the sex
is rising. These people who	(XVPR) come in $come to$ Belgium
Family planning	(XVPR) consists on $consists of$
have the possibility to	(XVPR) discuss about $discuss$ their problems
which the purchaser cannot	(XVPR) dispense of $dispense with$
the health. Nobody	(XVPR) doubts about $doubts$ that.
harvest they get is often	(XVPR) exported in $exported to$ countries
of advice on	(XNUC) a 0 better health care
for years. Undoubtedly	(XNUC) a 0 big progress has been made
characteristic	(XNUC) behaviours $behaviour$
It provides	(XNUC) employments $employment$
combining study life and	(XNUC) entertainments $entertainment$
are many other	(XNUC) leisures $leisure facilities$
a balance between work and	(XNUC) spare times $spare time$
need to do some	(XNUC) works $work$ or simply for your personal

It is also worth noting that, as pointed out in Granger (2002:14), there are two methods for carrying out CEA: in the first one, the analyst chooses an "an error-prone linguistic item" be it a phrase, a word category, a structure etc. and looks through the corpus for error instances in this item to retrieve them using computer retrieval software. In the second method, the analyst tags either the whole corpus or maybe a category such as auxiliaries, verb tense etc. for errors. The first method is obviously easier and less time-

consuming but the second one is more informative and useful (ibid).

4.4.1 The error taxonomy and computer tool used

As stated above, the aim of this research study is the measurement of accuracy and complexity in the essays written by Kurdish students. At the heart of measuring accuracy lies the issue of correct forms and errors in terms of number and severity (see Polio and Shea, 2014; Wolfe-Quintero *et al.* 1998; Polio, 1997). Therefore, an error taxonomy has to be used for describing and classifying the errors and a computer error tagger for tagging them. For this purpose, the Université Catholique de Louvain's taxonomy of error and Dexter Coder as a computer tool are used.

4.4.1.1 Université Catholique de Louvain's taxonomy

whether something is an error or not is not as difficult as under which category the error has to be subsumed. It is a decision that is subject to high levels of fuzziness and subjectivity i.e. intuitive decisions. If the error taxonomy, which the error analyst uses, is not 'consistent' enough (Granger, 2003 cited in Díaz-Negrillo and Domínguez, 2006:88) this will create inconsistency and similar errors or in some cases the same errors might be categorized differently by the different raters or sometimes by the same rater. I have chosen Université Catholique de Louvain's error taxonomy because it helps the rater to be consistent throughout the error description process. The errors are well classified linguistically and ambiguity and inconsistency. This taxonomy (see Dagneaux *et al.* 1998 for a detailed description), as shown in table 4.1 below and appendix G, is a hierarchical linguistic classification, which covers Formal, Grammatical, Lexical, Lexico-grammatical, Word, Sentence/style errors as main error categories. This covers almost all the classes under which an error might be subsumed including even the pragmatic or discourse level.

Another feature of Université Catholique de Louvain's error taxonomy is its granularity or delicacy. The main levels are well subdivided into more categories. For example, errors in grammar are further divided into errors in verbs, nouns, adjectives, adverbs etc. These subcategories are also further divided into sub-codes like tense, number, superlative/comparative and order respectively. So, an error might be classified as GVT (Grammar, Verb, Tense). The subcategories provided are detailed and comprehensive with examples and hence reduce the possibility of ambiguity. For example, the lexico-

grammatical category is further divided into nine well-defined subcategories which are clearly stated and differentiated allowing the analysts to choose from many and hence . Also, at the level of lexis six categories are provided: (1) Lexis, Conjunction, Subordination (LCS) for indicating any erroneous case of the use of subordinating conjunctions (2) Lexis, Conjunction, Coordination (LCC) for indicating any erroneous case of the use of coordinating conjunctions, (3) Lexis, Connector, Logical, Complex (LCLC) to describe errors in connectors like *in light of* , (4) Lexis, Connector, Logical, Single (LCLS) to describe errors in connectors like *nevertheless, moreover,* (5) Lexical Phrase errors (LP) to mark any erroneous case of the use of multiple word units, (6) Lexical Single errors (LP) to mark any erroneous case of single words like the case with wrong collocations. With all these categories the analyst will have a clearer idea of categorize an error.

As described above, the Louvain's error classification helps the analyst to be consistent in describing the errors and also it is detailed. It first describes the principles of error tagging and then exemplifies all the categories (see appendix G). Also, to borrow James' (1998:95) term, it is "user-friendly", hence less confusion when ambiguity arises. Moreover, it is flexible; new error tags can be added in case of occurrence of errors, which cannot be given any of the existing tags.

Table 4.1 Université Catholique de Louvain's error tag set (Dagneaux et al: 1996 cited in López: 2009:682)

CODE	TYPE OF ERROR
FM	form – morphology
FS	form- spelling
GA	grammar- articles
GN	grammar – nouns
GNC	grammar - noun case
GNN	grammar - noun number
GP	grammar – pronouns
GADJO	grammar - adjective order
GADJN	grammar - adjective number
GADJCS	grammar - comparative/superlative
GADVO	grammar - adverb order
GVN	grammar - verb number
GVM	grammar - verb morphology
GVNF	grammar - non-finite / finite verb forms
GVV	grammar - verb voice
GVT	grammar - verb tense
GVAUX	grammar – auxiliaries
GWC	grammar - word class
XADJO	lexico-grammar - erroneous complementation of adjectives
XCONJCO	lexico-grammar - erroneous complementation of conjunctions
XNCO	lexico-grammar - erroneous complementation of nouns
XPRCO	lexico-grammar - erroneous complementation of prepositions
XVCO	lexico-grammar - erroneous complementation of verbs
XADJPR	lexico-grammar - adjectives used with the wrong dependent preposition
XNPR	lexico-grammar - nouns used with the wrong dependent preposition
XVPR	lexico-grammar - verbs used with the wrong dependent preposition
XNUC	lexico-grammar - nouns: uncountable/countable
LS	lexis - lexical single
LSF	lexis - false friends
LP	lexis - lexical phrase
LCL	lexis - logical conectors
LCLS	lexis - single logical connector
LCLC	lexis - complex logical connector
LCC	lexis - coordinating conjunctions
LCS	lexis - subordinating conjunctions
WR	word redundant
WM	word missing
WO	word order
R	register
S	style
SI	style - incomplete
SU	style - unclear

4.4.1.2 Dexter coder

In this research, Dexter Coder was used because (1) it can easily be downloaded free of charge on its official website, www.dextercoder.org, (2) the codes can be highlighted with different colours, and (3) the tagged codes are retrievable in terms of frequency (the number of tokens of each error type) and concordances, where the erroneous structure can be seen in context. The number of tokens can be obtained directly on the Dexter Coder, but the concordances can be retrieved using Antconc.

Because Dexter Coder is based on XML, it cannot open normal MS word files. They have to be first saved as plain text files and then converted into XML files. As a result of this, Dexter Coder works in parallel with Dexter Convertor, which converts the plain text files into XML files. The following screenshots (Figures 4.4, 4.5, 4.6 and 4.7) show the whole process of first saving the file in the form of a plain text file and then converting it to an XML file using Dexter Convertor.

Figure 4.4 a screenshot showing saving the file as a plain text

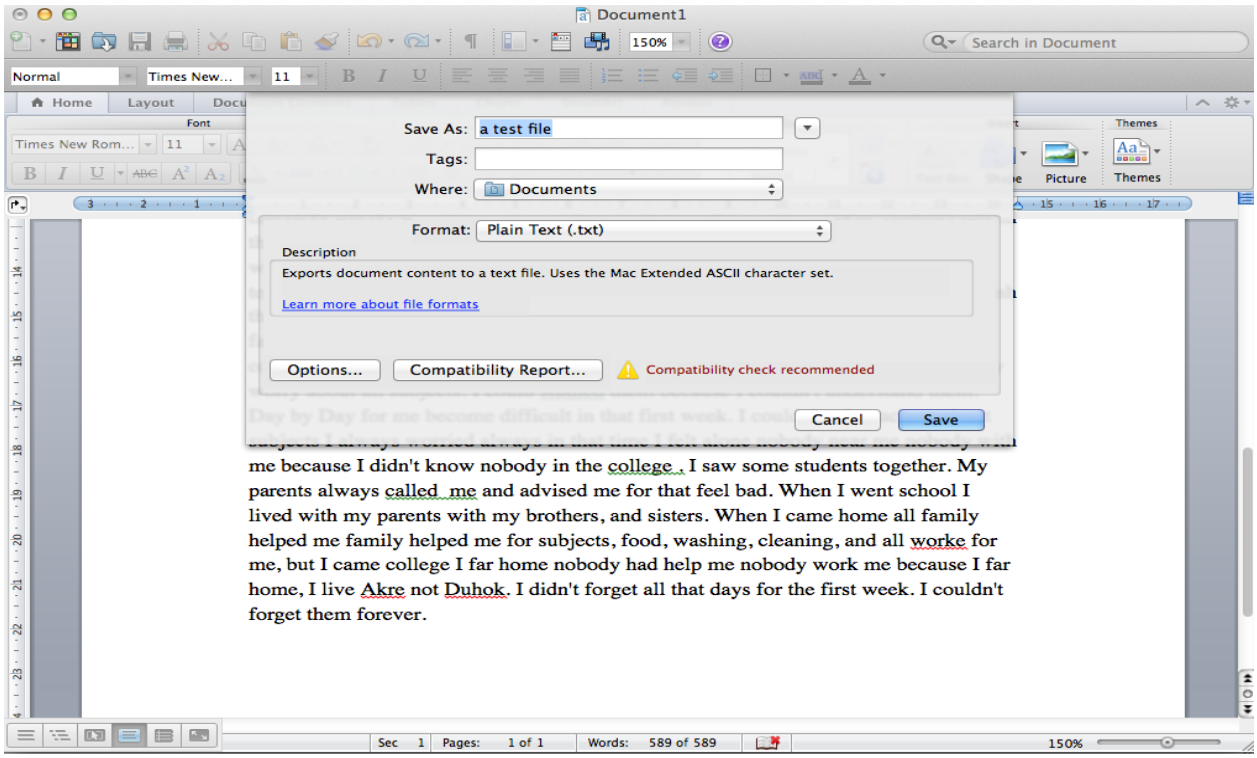


Figure 4.5 a screenshot further showing saving the file as a plain text

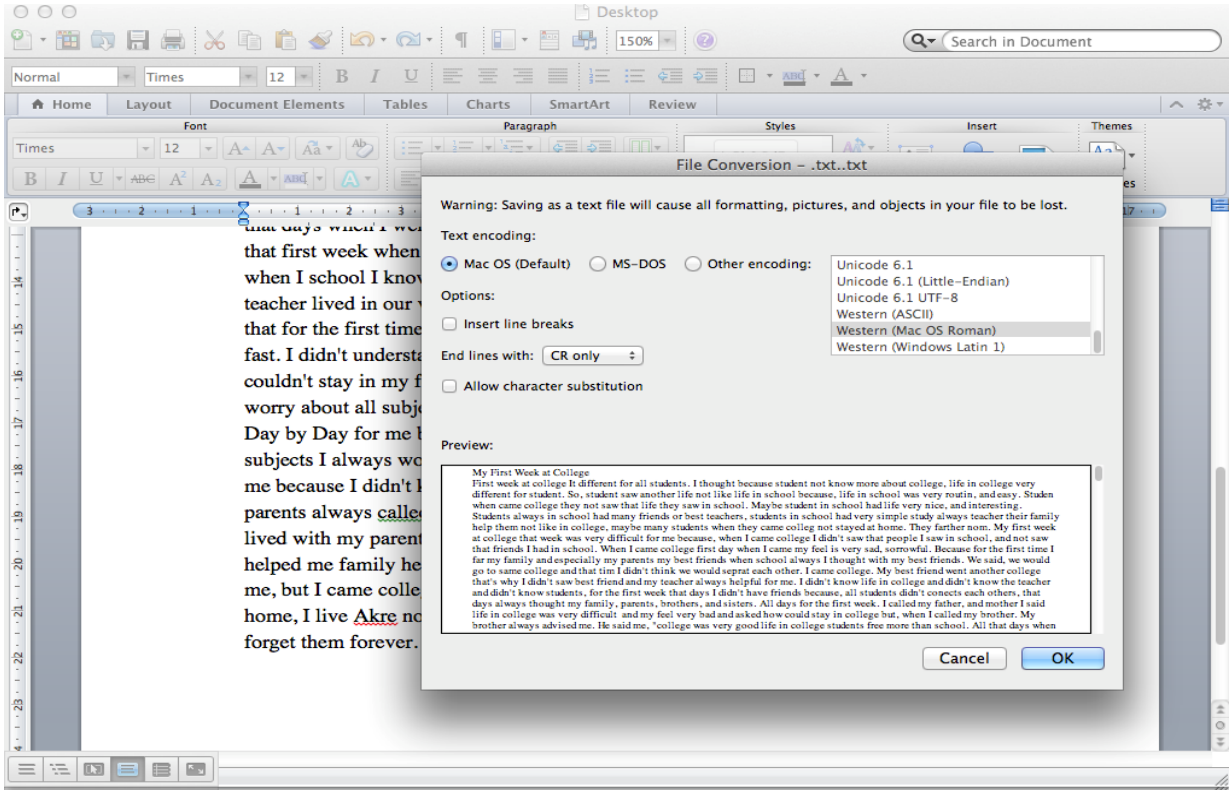


Figure 4.6 a screenshot showing opening the plain text file in the Dexter Convertor

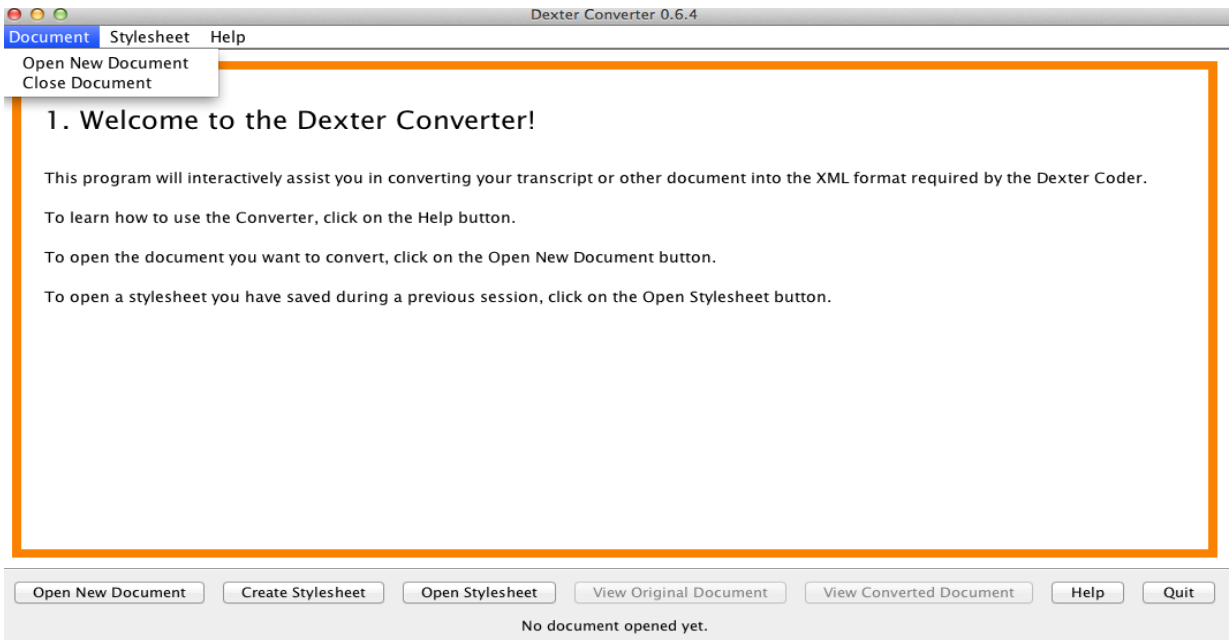
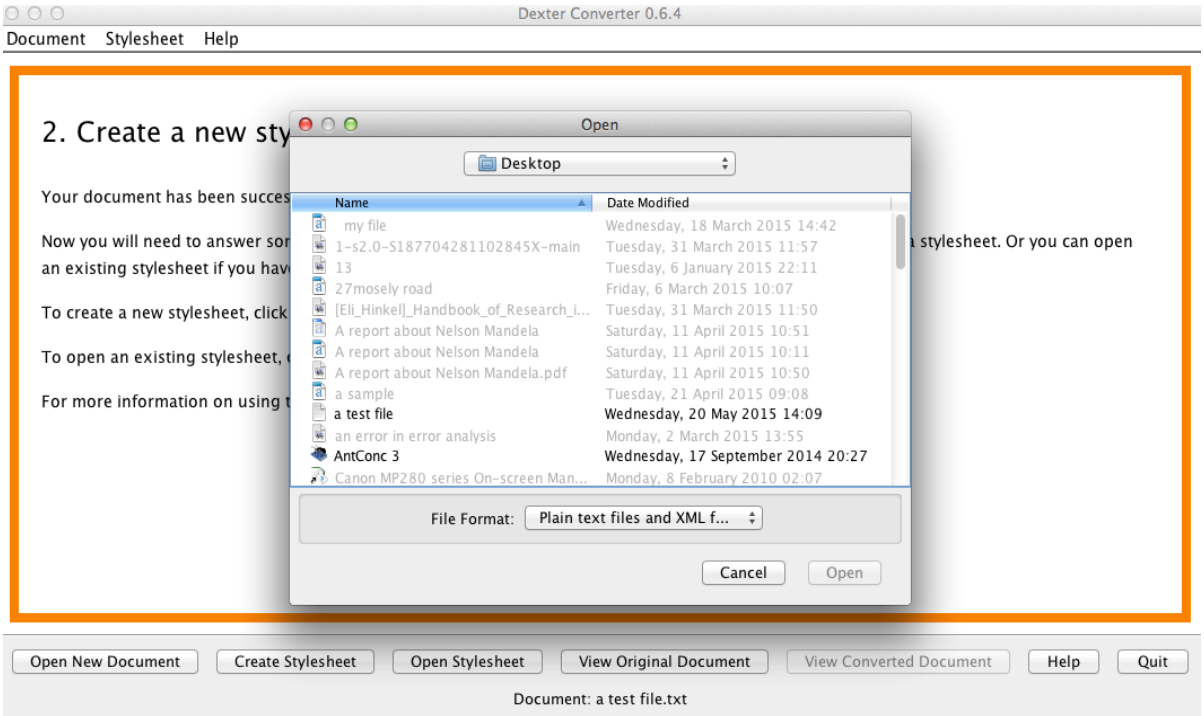


Figure 4.7 a screenshot further showing opening the plain text file in the Dexter Convertor



As these screenshots demonstrate, the plain text file is opened using the Dexter Converter.

However, to convert the file into XML a stylesheet has to be created according to the characteristics of the text, e.g. is the text spoken or written? and the mechanics used in the text like the quotation marks, parentheses etc. This stylesheet can be saved and applied later to other files as appropriate. The following screenshots (Figure 4.8 and 4.9) show the process of creating a stylesheet for a file. As soon as the user of Dexter Coder creates the appropriate stylesheet for their file, they can proceed with its conversion to XML file.

Figure 4.8 a screenshot showing the creation of a stylesheet for the file

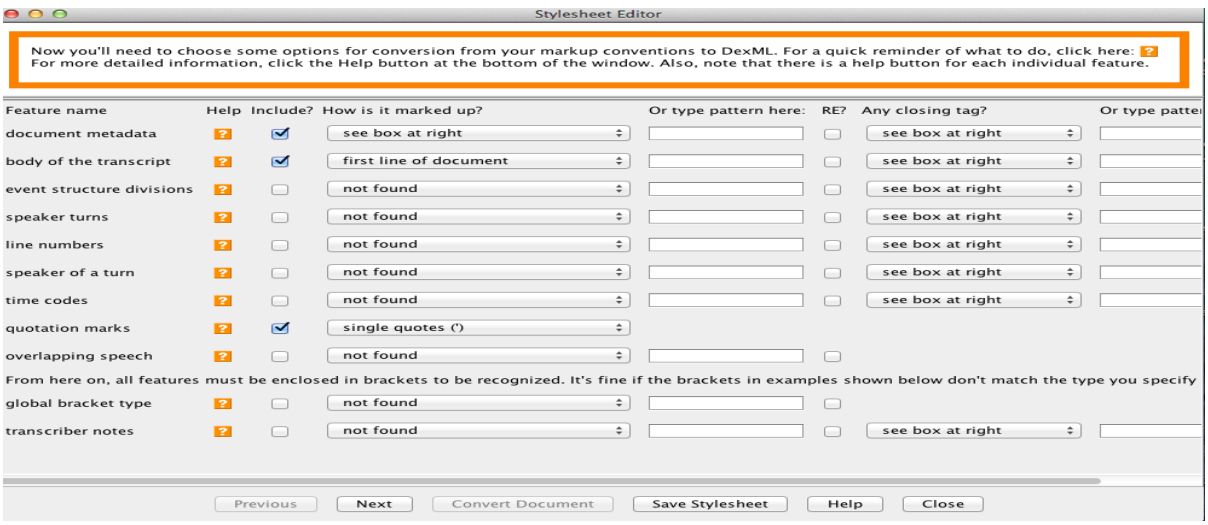
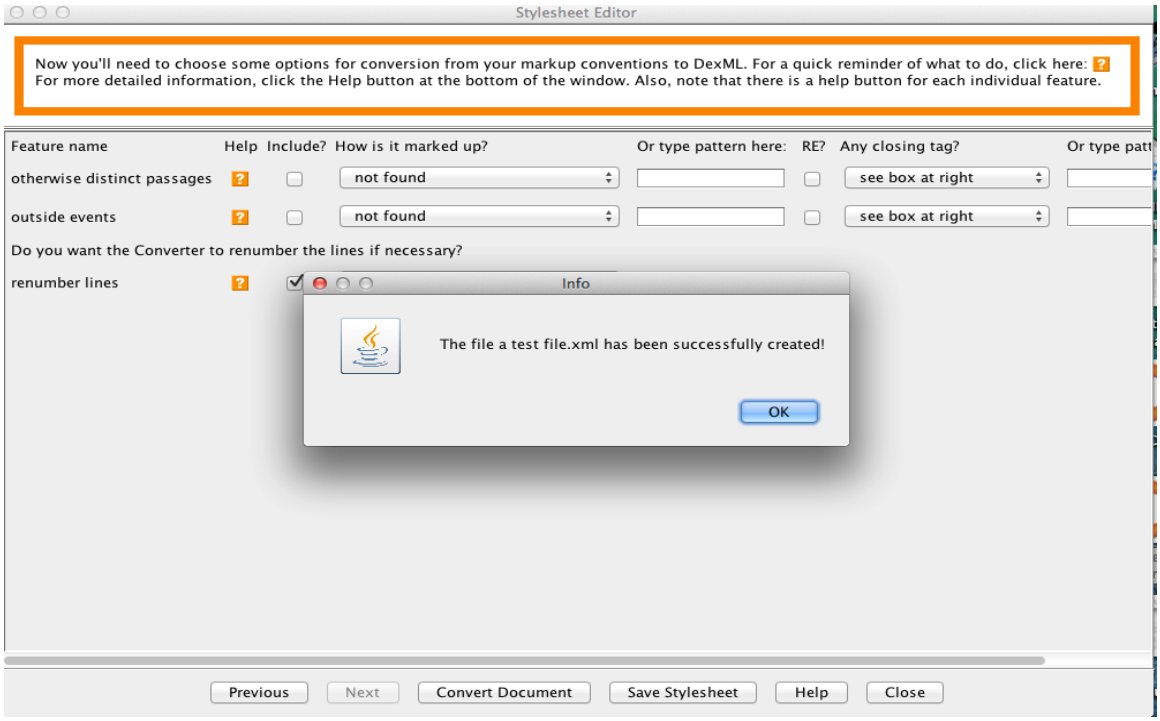


Figure 4.9 a screenshot further showing the creation of a stylesheet for the file



Now the saved file can be worked on with Dexter Coder as shown in figures (4.10, 4.11, 4.12). First, the XML file that the user has created can be opened using Dexter Coder. Then, codes can be created and named by clicking on the *new code*. Any part of the text can be coded with any created code by selecting the code and the part of the text to be coded and then pressing A on the keyboard.

Figure 4.10 a screenshot showing the creation of a new code with the name of Test Code (TC)

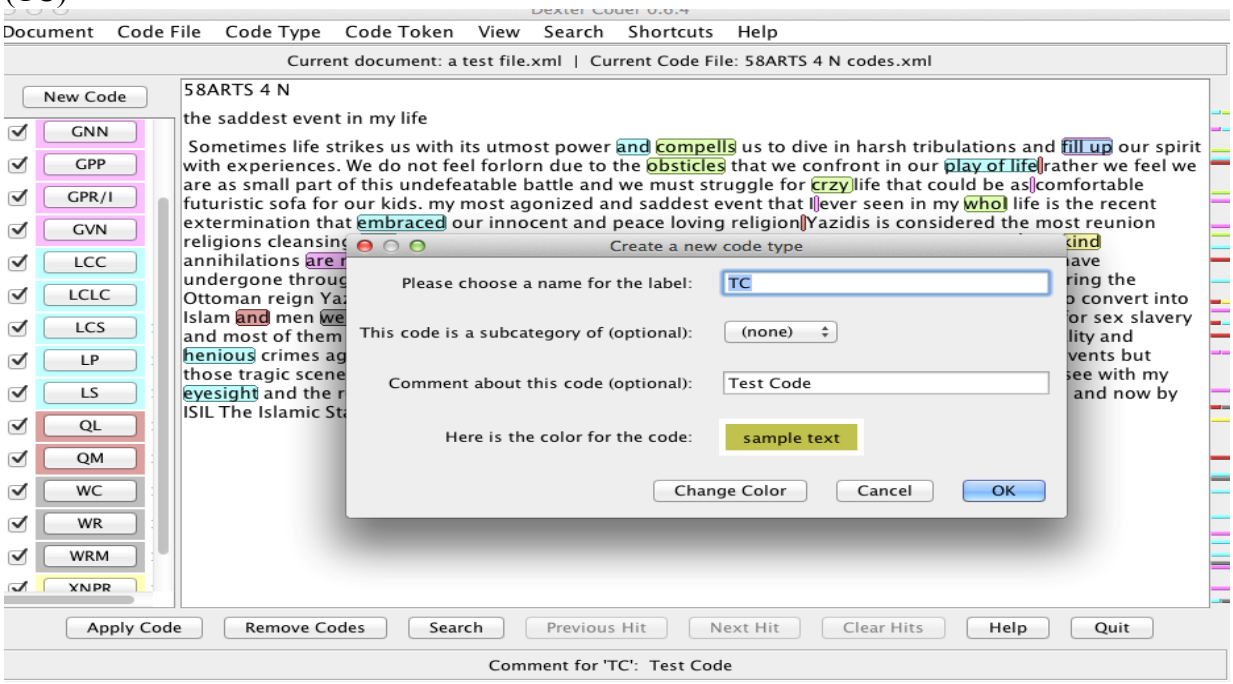


Figure 4.11 a screenshot further showing the creation of a new code with the name of Test Code (TC) in yellow color.

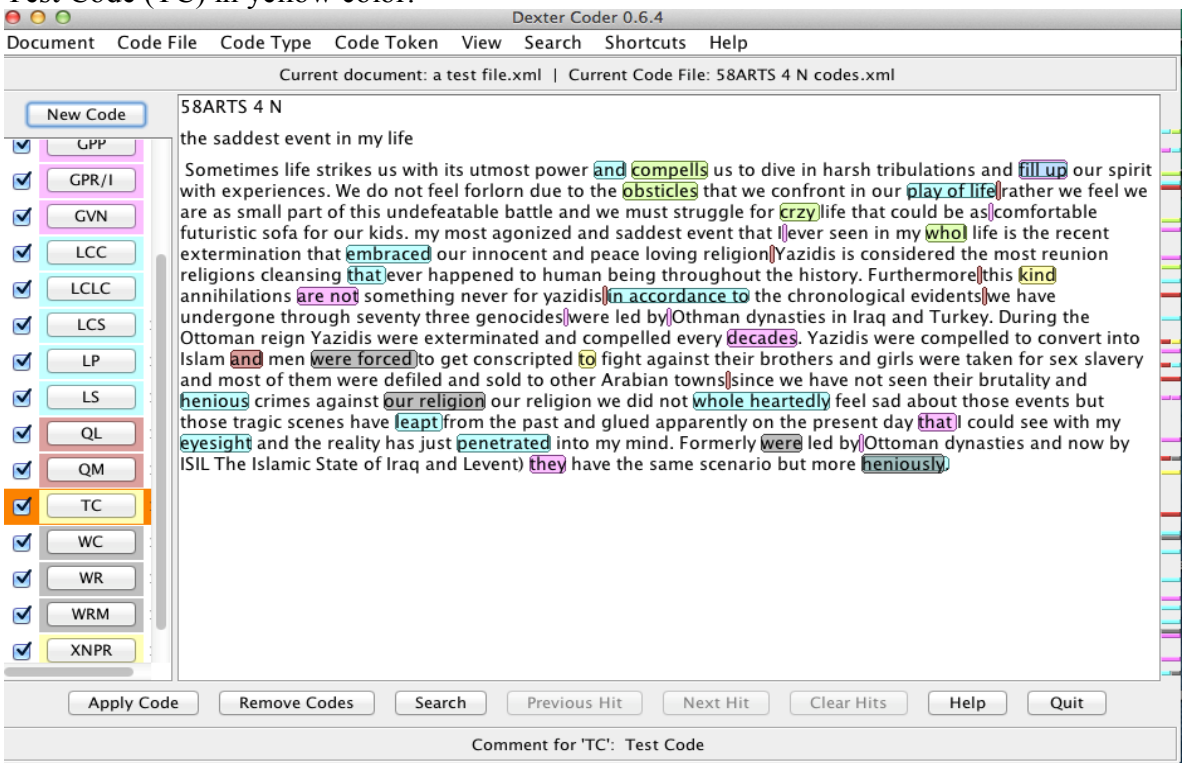
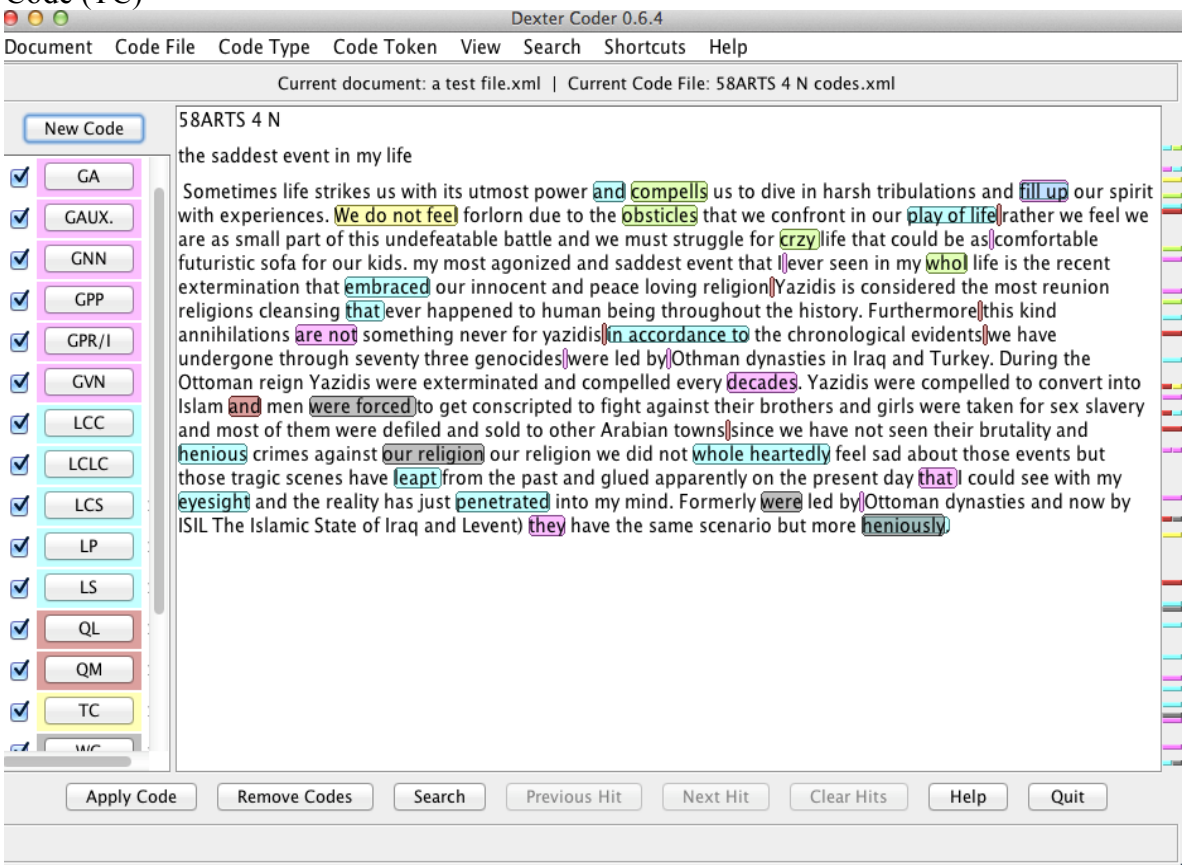


Figure 4.12 a screenshot showing the application of the new code with the name of Test Code (TC)

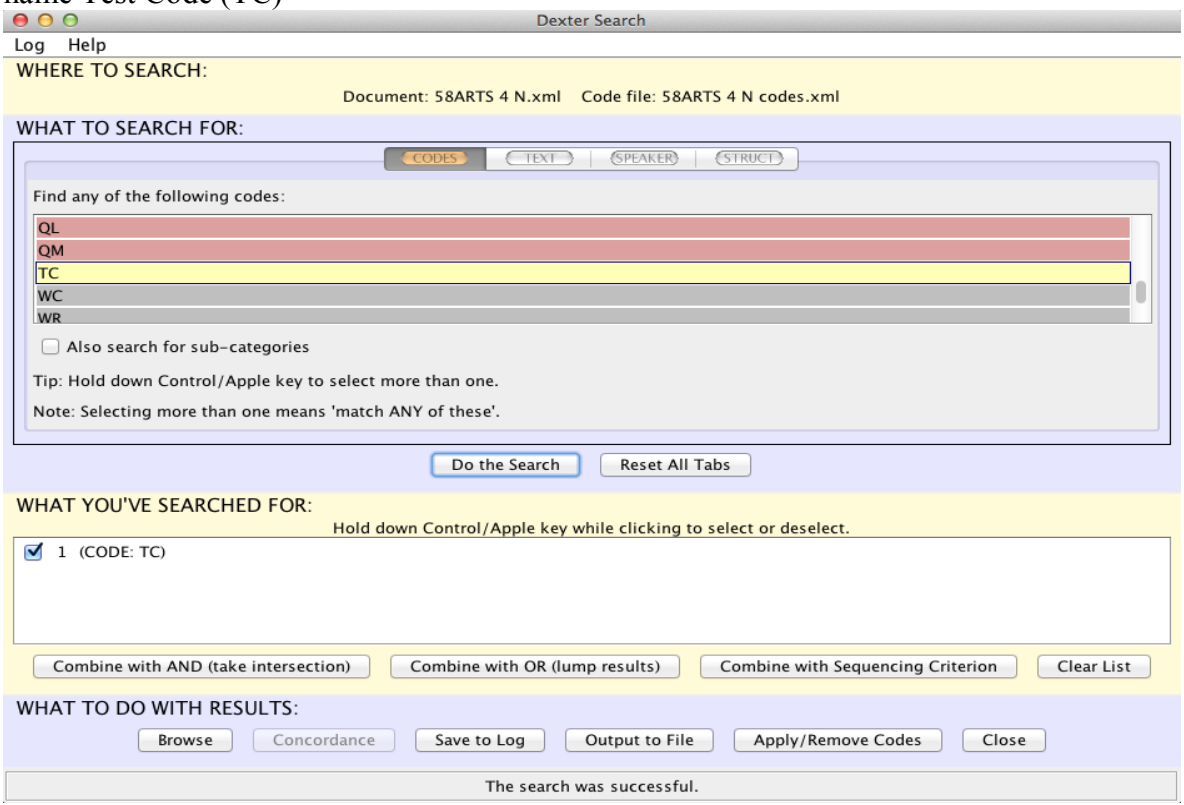


As appears in the screenshots, a code has been created with the name TC having a different colour from the other codes. To apply this code to the text, the code was selected first, the part *we do not feel* from the text was then selected and after that, the key A on the keyboard was pressed. Looking at the figure 4.12, one can see that the code TC is applied to the part of the text *we do not feel* which is in yellow color.

Another important point is that the user of the Dexter Coder can save the code file with the same name as its XML file (i.e. that contains the text) for easy reference. Then, for opening the code file, the user has to first open the XML file. What is more interesting is that one can have a number of code files for the same XML (text) file. Thus, one can work on the same file for different purposes. For example, the Dexter Coder user can code the errors once in the file and save that code file and later code the correct forms in the file and save that code file too.

The significant task that can be carried out by using Dexter Coder is the search for the number of tokens of each code in the file. It can therefore calculate the errors. The following screenshot (figure 4.13) shows the search for the code TC.

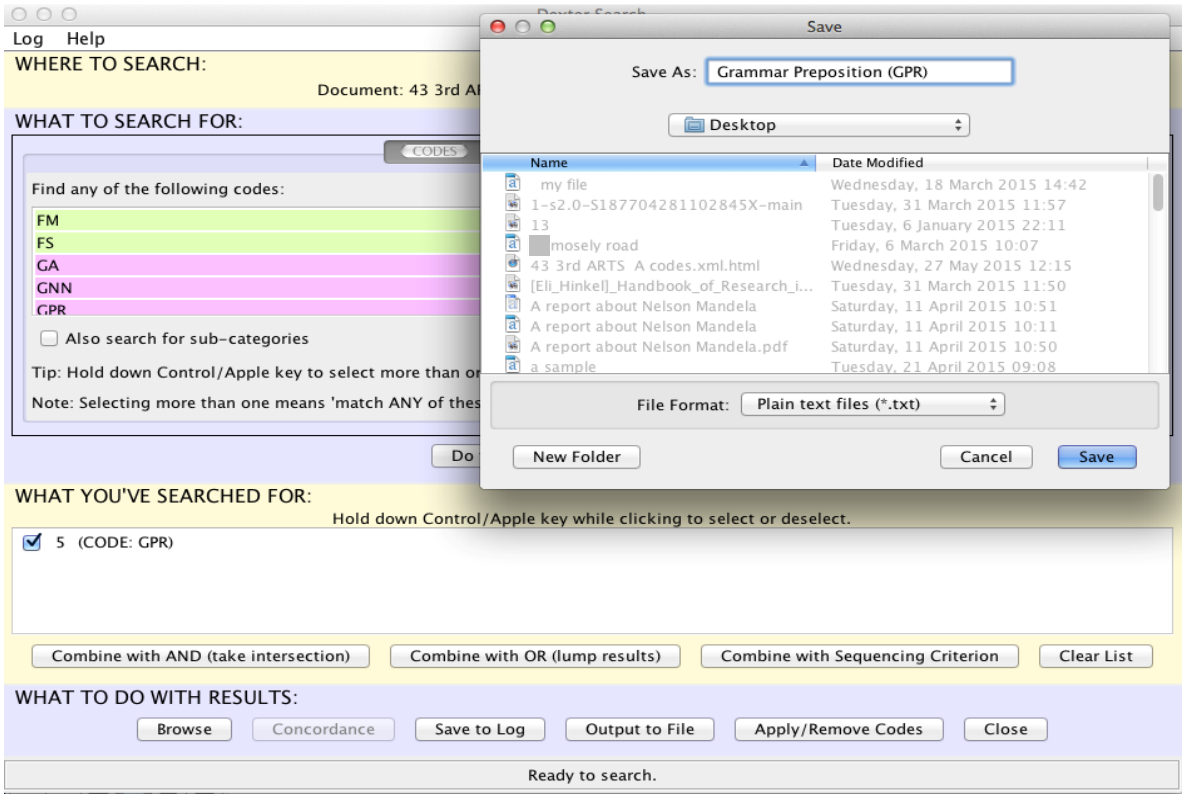
Figure 4.13 a screenshot showing the retrieval of the tokens of the new code with the name Test Code (TC)



It is evident that there is only one token of the code TC.

All the coded errors can be seen in context separately in the forms of concordances; this can be achieved by saving the code tokens as a plain text file and then opening them in the form of concordances in Antconc. This can be done by pressing the key *Output to File* in the search window and then saving the file as a plain text file. After that, the file can be opened on AntConc and the following regular expression: <<[^>]*>> can be applied to see the errors in context in the form of concordances. The following screenshots (Figure 4.14 and 4.15) display this process.

Figure 4.14 a screenshot showing saving the preposition errors in the form of a plain text file



This figure clearly demonstrates saving the error tag GPR (Grammar Preposition) in the form of plain text file to be opened later on AntConc.

Figure 4.15 a screenshot showing opening the error tag (GPR) in the form of concordances on AntConc 3.4.3m

As this figure shows there are five instances i.e. tokens of the error tag GPR. They are displayed in context.

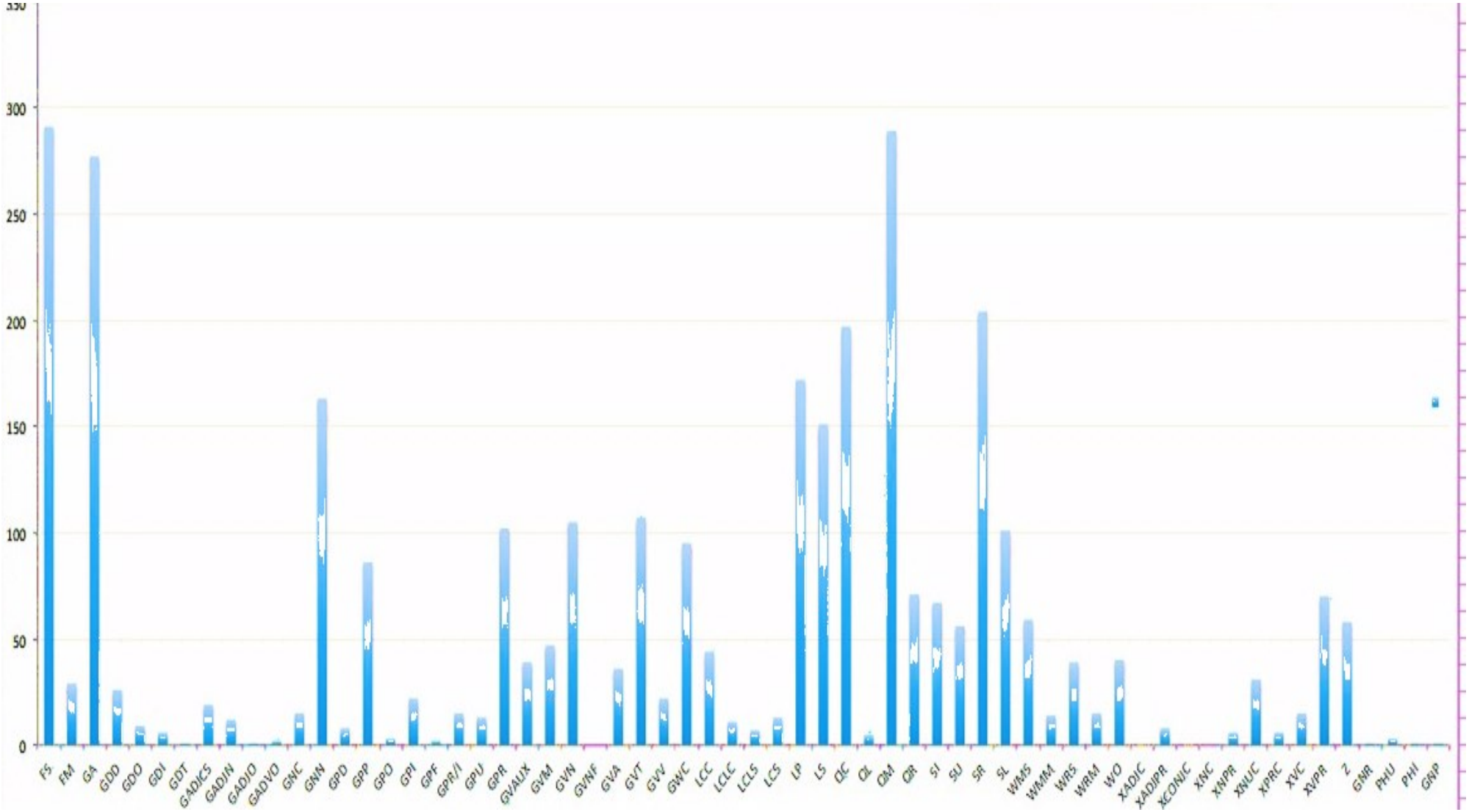
4.5 Error tagging of the data of the present study

A sample of the data of the present study (48 essays) has been subjected to error analysis using Université Catholique de Louvain's taxonomy and Dexter Coder for tagging the errors. However, the codes GPR (Grammar, Preposition), PHI (Phrase Incomplete), PHU (Phrase Unclear), GNP (Grammar, Negative particle), SL (Stretch, Ill-formed) and SR (Sentence, Run on) were added because errors that could not be classified according to de Louvain's taxonomy were found in the data. Although punctuation confusion could account for run-on sentences, I added this tag to show the actual rate of these sentences because as I analyzed the data, I found a great number of such sentences. In order to ensure a high rate of consistency in error tagging, the identification of error was done by the researcher (me) herself. The tagging was carried out by reconstructing the sentence into its correct form and then tagging the errors in it to simply adjust it to this form. Also, both appropriateness and acceptability are taken into consideration while identifying the errors.

4.6 Results and discussion

The following figure shows the number of errors in each code

Figure 4.16 number of errors in a sample of the present study data (no. of subjects=48)



As the figure 4.16 shows, the spelling (FS), articles (GA), and missing punctuation (QM) errors rated the highest in their share of the total number of errors[7]. Errors of wrong noun pluralization (GNN), punctuation confusion (QC) and run-on sentences[8], wrong lexical phrase (LP), wrong use of a single lexical item (LS), ill-formed stretches (SL), wrong use of prepositions (GPR), ungrammatical use of verb number (GVN) and tense (GVT), and erroneous use of word class (GWC) and personal pronouns (GPP) come next in their share of the total number of errors. Errors of redundant punctuation (QR), sentence incomplete (SI), sentence unclear (SU), verb followed by the wrong dependent preposition (XVPR) scored a little less than the errors that come second in their rate of the overall number of errors. The following are examples of these errors from my data (S=Subject):

1. Spelling (FS) e.g. 4.3 S(8) *I had anervious and anxsious feeling*) [9]
2. Articles (GA) e.g. 4.4 S(17) *if you are not good in English you may have problems with_objects[10] that you have)*
3. Punctuation missing (QM) e.g. 4.5 S(22) *oneday I am going to outside and my mother wasn't in home_I was a child__I am going to outside of home inorder to play with my friends__at that time my father also wasn't at home)*
4. Noun pluralization (GNN) (e.g. 4.6 S(12) *There are many purpose of having a journey and I will mention some of them)*
5. Punctuation confusion (QC) and run on sentences[11] e.g. 4.7 S(5) I *didn't know life in college and didn't know the teacher and didn't know students, for the first week that days I didn't have friends because, all students didn't conects each others, that days always thought my family, parents, brothers, and sisters*).
6. Wrong lexical phrase (LP) e.g. 4.8 S(29) *that they leave our brains without asking for permission*),
7. Wrong use of a single lexical item (LS) (e.g. 4.9 S(3) *After she died I returned like a child who don't know what he is doing*).
8. Stretch ill-formed (SL) e.g. 4.10 S(15) *so many time when we are intrest with some thing make us forget that thing make us forget how to laugh and there are many thing you have to do, because to continue your funy life.* e.g. 4.11 S(49) *We help each other in make forget people, beacuase, there very crowd and tired work*
9. Wrong use of prepositions (GPR) e.g. 4.12 S(34) people work at farm)
10. Ungrammatical use of verb number (GVN) (e.g. 4.13 S(12) *Finally, travelling are goods for our minds and body*) and tense (GVT) (e.g. 4.14 S(4) *After we got money we take him to a very good doctor in sham and after five months my father recovered and was a glad)*

11. Erroneous use of word class (GWC) e.g. 4.15 S(23) *I will try to speak in a relax*) and personal pronouns (GPP) e.g. 4.16 S(19) *Istanbul had a nice and beauty nature like sea, bridge of Muhamad Fatih and many nice mosque that I didn't see them in my country*
12. Errors of redundant punctuation (QR) e.g. 4.17 S(25) *When I first met him and knew him I was surprised, because I always said it is impossible for me to have a friend like him , because those days we can not trust other people, they always betray each other.*
13. Sentence incomplete (SI) e.g. 4.18 S(5) *When I came college first day when I came my feel is very sad, sorrowful. Because for the first time I far my family and especially my parents my best friends*)
14. Sentence unclear (SU) e.g. 4.19 S(16) *so I can say first week was as place that you visit and have any body there you don't anyone at first week*).
15. Verb followed by the wrong dependent preposition (XVPR) e.g. 4.20 S(21) *When I came back to my relative my poor father was very very apset about me because he didn't know what will happened for me).*

The first three types of errors that rated the highest seem to fall into completely different categories. For example, spelling is a form error, punctuation is an error in mechanics and errors in the use of article are grammatical errors. This shows a defect in subjects' English on three different levels. Research has proved that the system of English articles is difficult for non-native speakers to acquire (see Miller, 2005). Although the primary cause of this difficulty is "unclear" (Butler, 2002: 451), it is suggested that "a number of conceptual differences with regard to their [learners'] considerations of the hearer's knowledge, specific reference, and countability, … may account for learners' errors in article use across different proficiency groups" (ibid: 451–452). The difficulty the subjects face in spelling might be due to the fact that Kurdish, the mother tongue of the majority of them, is spelt phonetically while English is not. Following the spelling system of English, memorization plays an important role and this increases the likelihood of producing incorrectly spelt words. Concerning punctuation, although it is taught in high schools and reviewed by many teachers at university, the area remains difficult for Kurdish students. This difficulty might be attributed to two reasons. Firstly, the subjects wrote about one narrative topic, a story-like topic; thus, they sequenced the events one after the other and because their knowledge of lexical items denoting sequence might be limited, they resorted to either no punctuation or the use of a comma, thinking that this will express their ideas in sequence. Secondly, the subjects have mostly used a spoken style of writing, except for a small number of them who adhered to the principles of writing, which led to them place sentences one after the other without any

punctuation or with a comma, again guessing that a comma joins two sentences conveying equivalent or similar ideas. This spoken style might have been learned as a result of their excessive use of the Internet or watching English movies, for example.

Moreover, what also attracts attention among the second group of errors is the ill-formed stretches (I will call them stretches because they cannot be identified as only sentences or clauses whereas a stretch might be a sentence, clause, phrase or in some cases a whole paragraph), which comprise language that is too ill-formed to be reconstructed into a correct version. All these stretches were indeed so problematic that they could not be analyzed for errors. Consider the following example:

e.g. 4.21 S(24)

I say something about my day when I saw or met my friend and also I will descripe inner beauty and outer beauty about my friend but firstfull:- I will say about the day was a very nice day was a sunshine and beautifully day and I was a relax in day and happy so any how she came in the beautiful day.

Not to leave these stretches unanalyzed or OUT_3 as per Corder's algorithmic error analysis model in figure 4.1 above, the whole stretch was marked as ill-formed. These stretches can be regarded as good evidence that sometimes learner language is too erroneous and therefore resistant to error analysis.

Another interesting case is unclear sentences. Sometimes the whole sentence was so incomprehensible and could not be understood even by referring to the wider context (the whole essay). Most of this incomprehensibility was due to the wrong lexical items used in the sentence. Example 4.22 shows one of these cases:

e.g. 4.22 S(6)

to show that how much good that someone who is always willing to hoist his honour and behavior and to help any person in his life, but it depends on persons

There were some cases where the sentence was both ill-formed and unclear, so it was marked as both.

Also, very common in the data were run-on sentences (or these can also be called stretches because they are different sequences of language) that were placed together with either no punctuation or with a confused punctuation, mostly commas. Or, they were combined with a redundant coordination. Example 4.7 in point 5 above exemplifies this error.

Likewise, the excessive number of wrong lexical phrases and lexical single words, as shown in the examples 4.8 and 4.9 in points 6 and 7 above, is noteworthy. This is probably due to L1 interference because if anyone who has knowledge of the Kurdish language checks the two sentences, they will be able to directly show their exact equivalent translations in Kurdish. What is unexpected here is that these cases existed even in the good quality essays where incomprehensible and/or ill-formed sentences were rarely found. This can be an indication of the fact that even where the subjects' grammatical or syntactic accuracy improved, there still remained shadows of L1 interference exhibiting itself in the phraseological patterns. One point one can possibly conclude from this is that their lexico-grammatical accuracy is still affected by their L1 even after the syntactic rules have been learned (or acquired). Perhaps, this can bear witness to the point that phraseological patterns are acquired late or it might be a sign of fossilization[12].

As explained above, the types of errors are grouped into three groups; the first group that rated the highest, the second group that scored less in the total number of errors and the third group, which although it scored a significant number of errors from the overall number, was less than the second group. Looking at these three groups, it is evident that mostly the second and the third group of errors seem to have rendered many of these essays almost incomprehensible and of bad quality writing. Only punctuation missing (QM) in the first group accounted for some of the incomprehensibility but it is the lexical errors (LS and LP), ill-formed stretches (SL) and unclear sentences (SU), punctuation confusion (QC) and incomplete sentences (SI) (i.e. fragments) that caused the highest degree of incomprehensibility in the data. For example, it is the contradiction between the words *place (-time)* and *week* (+ time) in e.g. 4.19 in point 14 above that has caused incomprehensibility not the missing definite article that should have preceded the word *first* to make a complete and well-formed noun phrase *the first week.* This is one of the lexical examples that caused the whole sentence to be unclear. Also, examples 4.10 and 4.11 in point 8 show how the ill-formed stretches rendered almost a whole paragraph ill -formed and almost incomprehensible.

4.7 User engagement

Section 3.6 in chapter three discusses the method of user engagement, where 20% of the data which were analyzed for each chapter were given to another rater (or teacher who will very likely use the research) to reanalyze in a discussion with me as the first rater. Accordingly, in order to ensure that the error tagging used for this part of the study is less subject to the researcher's intuitive decision, the coding of 20% of the data that had already been tagged by

me as the researcher has been reviewed by another teacher specialized in linguistics. A second rater who has a background in applied linguistics has reviewed the whole process of tagging the data using Université Catholique de Louvain's taxonomy and basing her analysis on its detailed manual, which exemplifies each tag. The discussion revealed a number of important points:

1. Clear cases of punctuation did not lead to a lot of discussion. However, the case whether it is obligatory to put a comma before a subordinating or coordinating conjunction raised a real discussion e.g. 4.23 S(10) *We had a very interesting week, because we went to places that I had never seen them before* . I have considered the comma that precedes *because* as an error but the second rater argued that, on the contrary, not putting this comma is an erroneous case. Other examples of punctuation that provoked discussion were instances like 4.24 S(3) *Before 4 days of our marrige we was on the way to go to Arbil and buying some clothes and other things there.* The second rater suggested that a comma separate *Before 4 days of our marrige* from the rest of the sentence.
2. The identification of most of the errors was not difficult. We both highlighted almost the same cases as erroneous. Their description (labeling), however, was arguable. An example of this is the sentence produced by the subject (2) 4.25 *Also I tried to make my relationship with my teachers as good as I could because they would be my right hand for encourage me and develope my language from the worst to the best*. We both agreed that *encourage* is an error. I labeled it as a XPRCO, which means a lexico-grammatical error of preposition with the wrong complementation. The second rater labeled it as Grammar, Verb, Morphology (GVM) supporting her decision by the point that the existing verb must end with *–ing,* which is a morphological error. Another example was produced by (e.g. 4.26) the subject (4) *After his recover*. I have labeled *recover* as an error in word class but the second rater stated that alternatively it could also be marked as an error in spelling.
3. Sometimes labeling the errors was difficult and controversial because of the way correction is done. We both agreed that example 4.27 S(2)*The result I became the person who I hoped to be* is erroneous. I corrected the sentence in the following way *The result was that I became the person who I hoped to.* However, the second rater corrected it to *As a result of that, I became the person who I hoped to be.* In this case the coding of the same error differs from one rater to another. Also, another example (e.g. 4.28) that was produced by the subject number (6) *That friend that I have, he is always enjoying with me.* I considered *that friend that I have* as a fragment and did

not consider *he* as a redundant pronoun. The second rater corrected the sentence to *the friend that I have always enjoys time with me*. With this second rater *he* is considered an error of pronoun (GPP).

4.8 Conclusion

This chapter has examined one method for measuring accuracy in a number of the essays considered for the present study (48 essays), which is the identification and description of errors in whole samples. The errors are first identified and described and then grouped into three groups i.e. errors group 1, errors group 2, errors group 3. The first group, which rated the highest, included errors in articles, errors of missing punctuation, and errors in spelling. These three types of errors seem to fall into different categories affecting different levels of language. Errors of articles are grammatical errors, errors of missing punctuation are errors in mechanics and spelling errors are errors in form. The second group included errors like errors of wrong noun pluralization, punctuation confusion and run on sentences, wrong lexical phrase, ill-formed stretches, wrong use of prepositions etc., and the third group included errors of redundant punctuation, sentence incomplete, sentence unclear etc. 20% of these 48 essays have been reanalyzed by another rater in a discussion with me and many points were discussed.

It is important to highlight that the essays contained a lot of global errors that could not be easily identified. In other words, the language of many essays had to be rewritten as it contained many ill-formed stretches and unclear stretches that could not be analyzed with error analysis. It is also quite obvious that the method has left the correct language unanalyzed which comprises a considerable part of the language in the case of good quality writing. One of the solutions for these two issues is trying the other facet of analysis i.e. the codification of the correct parts of the language rather than the erroneous parts, and this is the core of analysis and discussion in the next chapter. I have concentrated on the number of errors in this chapter but in the next chapters I will compare individual essays and also compare the results of the third year students to the results of the fourth year students.

Notes

1. The concept of complexity is also dealt with in chapters eight and nine.
2. This method has been described in detail in chapter five
3. This term and other terms like the "idiosyncratic dialect" are used by Corder (1981) to refer to learner language. Other scholars use different terms. For example, Selinker (1972) uses "interlanguage" and Nemser (1971) "approximate system". All these terms imply that learner language is a system in itself.
4. Corder's paper 'Idiosyncratic dialects and error analysis' was also previously published in 1971 in the *International Review of Applied Linguistics* and in 1973 in Svartvik, J. (ed.).
5. The algorithm with its questions marked with alphabets is the idea of Brown (2007).
6. See also the example given by Schachter and Celce-Murcia (1977:445).
7. A table showing the number of errors is attached as appendix H.
8. Which reflects punctuation confusion and punctuation missing.
9. Underlining indicates an error.
10. Missing the definite article *the*.
11. One of the codes that was added to the Université Catholique de Louvain's error taxonomy was a run on sentence.
12. See Han (2004) for the definition of fossilization.

Chapter Five

Correctness Analysis (CoA)[1]

5.1 Introduction

In the previous chapter, the method of error analysis has been applied to a sample of the data; however, previous investigations have focused on measuring correctness rather than error, calculating the accuracy of units such as clauses and T-units. In this chapter, I am going to describe these units and measures and apply them to a sample of the present study. Since, in these measures correct units are considered, I have called this method *correctness analysis* (CoA) as opposite to *error analysis*. For the most part, I will be replicating what other researchers have done but towards the end of the chapter, I shall propose a new way of measuring accuracy.

Based on what is stated above, this chapter is intended to answer the following questions: (1) Which units have been used as baseline units in the previous studies of CoA and how did researchers deal with the specific characteristics of the texts they investigated? (2) What types of measures are used in previous research that measured accuracy in language with focus on those that assessed correctness rather than error? (3) Can a number of these measures of correctness be usefully applied to the data I collected for the present study? (4) If so, to what extent are the measures applied useful in terms of differentiating between levels of students, assessing the English the subjects have produced without simply considering all their writing achievement as incorrect?

Thus, the theoretical part of the chapter begins with a detailed explanation of the various units reported in the literature for the study of correctness and a number of the studies that have used the method of correctness analysis. The practical part of the chapter describes the difficulties faced while assessing correctness in the corpus of this study. It also presents the results of applying the method of CoA to a sample of 80 essays of the data of the present study (third and fourth year students of two schools, 20 essays per each level per school). The units used for analysis are clauses and T-units and the measurement methods are the number of error-free clauses divided by the total number of clauses (EFC/C), the number of error-free T-units divided by the total number of T-units (EFT/T), and the number of error-free clauses divided by the total number of T-units (EFC/T). The chapter concludes with a

recommendation of adopting a new method of correctness analysis, which considers all units that are error-free regardless of their type i.e. they might be clauses, phrases, or sentences.

5.2 CoA as a measure of accuracy

CoA has been applied widely in language learning research aiming at measuring accuracy. Various units have been used in conducting this type of research, mostly depending on the type of data whether written or spoken. Below is a description of these units and a number of the research studies that have used ratio measures for the study of correctness.

5.2.1 The units used in measuring the accuracy of a text

The following are some of the units that are used in the measurement of accuracy.

5.2.1.1 T-units

The first unit is called the T-unit. T-units or "minimal terminable units" or the "shortest allowable sentences" were originally identified by Hunt (1966:737, see also Knoch, 2009 for a discussion on this topic). They are described as *minimal* because they are "the shortest units into which a piece of discourse can be cut without leaving any sentence fragments as residue" and they are described as *terminal* based on the fact that "it is grammatically acceptable to terminate each one with a capital letter at the beginning and a period or a question mark at the end" (Hunt, ibid). Hunt (ibid) further defines a T-unit syntactically as a "one main clause plus whatever subordinate clauses… [are] attached to that main clause". The following example is provided by Hunt and cut into T-units as follows:

e.g. 5.1

(T1) *I like the movie we saw about Moby Dick the white whale* ||(T2) *the captain said if you can kill the white whale Moby Dick I will give this gold to the one that can do it* || (T3) *and it is worth sixteen dollars* || (T4) *they tried and tried* || (T5) *but while they were trying they killed a whale and used the oil for the lamps*|| (T6) *they almost caught the white whale* ||

T-unit (5), for example, consists of one main clause: *they killed a whale and used the oil for the lamps* and one subordinate clause: *while they were trying.*

Here is another example from the data of the present study produced by the subject number 184:

e.g. 5.2 S(184)

(T1) *Healthy is the most happfuly things to make a best life* ||. (T2) *Good healthy gives us the greate key of clearly futures* ||, (T3) *so human's depending on his/her environment to be a truth of life* ||. (T4) *As well as life will be change by the places as human's choice*, ||(T5) *but most of people say; healthy is a part of environments* ||. (T6) *I do agree that the countryside life is much healthier than the city life because, countryside has much way to healthy like, environment place and crowded by other ways//*

This example (5.2) consists of six T-units. Notice that T-unit 4 is separated from T-unit 5 by the use of *but*. Notice also that T-unit 6 consists of one main clause *I do agree that the countryside life is much healthier than the city life* and one subordinate clause *because, countryside has much way to healthy like, environment place and crowded by other ways*

An orthographic sentence might consist of a number of T-units combined by a comma, or a coordinating conjunction, or nothing. That is why Hunt (ibid) calls the average number of T-units per sentence "the main clause coordination index" or the "sentence coordination index"

Consider the following examples from the data of the present study

e.g. 5.3 S(25)

(T1) *My friend always advised me, //* (T2) *he was always teaching me things, //* (T3) *he never let me alone, //(*T4) *he was their every time I needed him.*

e.g. 5.4 S(8)

(T1) *Third, students are very ohnest and kind*, // (T2) *and I became like a friend with them esaly,* // (T3) *and my friends in college are very good for me, in the class and outside class*, // (T4) *in the same time we had a very beautiful occuation together.*

e.g. 5.5 S(16)

(T1) *When I was a school boy I have said many times my father that why I have to go to school and stady //* (T2) *it is unuseful //* (T3) *and there are thing that are more important and interested works to do than studing //* (T4) *my father said to me if you dont go to school you cant reach the college and get what ever you want//* (T5) *so if you dont go to school there will be no college and no fun.*

In example 5.3, the sentence consists of four T-units separated only by commas and in example 5.4, the orthographic sentence is made up of four T-units separated by the coordinator *and* and also by commas, while in example 5.5, the sentence comprises five T-units separated neither by commas nor coordinating conjunctions.

Before applying this method to 80 essays in the data I collected for my study, a word is in

order about the explanation provided by Hunt (ibid) about T-units, i.e. "the minimal terminal units" and the syntactic definition. They do not seem to be compatible with each other. The first one is both meaning- and syntax-based (considering discourse) and the meaning-based part emphasizes the inclusion of fragments in one way or another so that they are not left as *residues* while the second (the definition) is only syntax-based and makes no reference to fragments. It is not only fragments that should be included in the definition but also other cases that may well be data specific. This is why a clear list of criteria that fits a given set of data will possibly help in removing any discrepancies and doubt in dividing the text into units and deciding about the presence or absence of error in that unit.

Despite the fact that many researchers have used T-units as an index of first and second language development, some points of criticism have been highlighted in Gaies (1980). The first criticism was raised by Ney (1966 cited in Gaies, 1980), who argued that T-units do not account for the "excessive coordination" within sentences (e.g. between noun and verb phrases, and clauses), even though they can account for the coordination between sentences. Consider these two examples by Ney (1966 cited in Gaies, 1980:55).

e.g. 5.6

Then the rain falls and spring comes. (2 T-units)

e.g. 5.7

So he went the woods and pulled the feather out of his hat from the partridges And put a flower in his hat and walked on through the woods. (1 T-unit)

What Ney (1966) argues against as noted by Gaies (1980) is the "arbitrariness" of the decision to include coordinated phrases in one T-unit and exclude coordinated independent clauses.

In the process of tabulating the data of the present study into T-units, I have faced an issue that is worth noting. When, in a complex sentence, two or more clauses are coordinated after a subordinating conjunction and also when two or more coordinated independent clauses co-exist in one complex sentence, will these be regarded as two separate T-units or one T-unit[2]?

e.g. 5.8 *While I was eating and he was playing, she came in.*
e.g. 5.9 *While she was eating, they came in and he greeted them.*

In addition to these two points of criticism, Gaies (1980) also raises the question of the extent

to which error-free T-units are sufficient to measure accuracy or there should be considerations of communicative appropriateness as well. But Vann (1978 quoted in Gaies, 1980:55) suggested that an error-free T-unit is one which "makes sense in [a] given context and [is] free both from morphosyntactic and lexical errors".

It seems that these decisions in cutting up the text into T-units (whether to include or exclude coordinated phrases and coordinated clauses in a T-unit) will most probably affect the process of tabulation because it will have an effect on the resulting number of T-units. Also, reaching a consensus as to what freedom from error means and the inclusion or exclusion of communicative appropriateness in that decision has an impact on this process.

5.2.1.2 Sentences

The second unit used in measuring accuracy is sentence. A sentence is defined by Homburg (1984:91) "as a string of words with a capital letter at the beginning of the first word and a period or another terminal punctuation mark after the last word". This definition poses problems for dividing the text in the data of the present study into sentences because a sentence might sometimes mean a whole paragraph where the writer continues juxtaposing strings of language without any sentence end marker. The following paragraph by S18 exemplifies this:

e.g. 5.10 S(18)

The day I met my best friend we talked about that days we past without see each others it was very hard for me and my friend, she told me every thing that happened for her about those days, and I told her every thing my happy days and my sad days like her our problems about that year my friend is chiman she is study phizech now in Duhok she is 4th year she is very kind, very beautiful and calm she is never lie with me and betry me, because she is love me and I love her so match so match I put her my best sister my best friend I cann't live without her my friend helped me last year I have a proble she can help me and she always with me in my hard time she could make me happy and forget my broblem and laugh and still now when I remember that problem she came to me and ask to me not talk about it and always consulate me and tell me that is no thing don't remember yourself, and another help when I came to English college I told her I don't want like English and I always sad because I didn't like it I want to go ingeneering college, but she is help me to stady her and stady English and I am happy now I am in this college he always told me English is ver good college, and has a good future and told me you can be agood teacher and transliter so I cann't how describe her because she is help me in all my life.

However, Bardovi-Harlig (1992:391) argues for the use of sentences instead of T-units. This may be attributed to the fact, as pointed out by Bardovi-Harlig (ibid) that dividing the text up into T-units might be an "artificial" imposition on the "psychological reality" of the sentence

as a unit. This is because the sentence has been produced by the writer as a unified "psycholinguistic" unit.

5.2.1.3 Clauses

A clause is defined by Polio (1997:139; see also Ishikawa, 1995) as a string that has "an overt subject and a finite verb". She excludes the presence of subject as a condition in the case of imperative. She does not consider a string with a subject and only an auxiliary as a clause giving the example e.g. 5.11 *John likes to ski and Mary does too* and claims that the underlined string is not to be regarded as a clause.

5.2.1.4 C-units and AS-units

As is true of the T-unit, sentences and clauses are mostly used in writing. The unit used in the study of speech is called 'C-unit' or 'communication unit'. A C-unit is defined by Foster and Skehan (1996:310) as "each independent utterance providing referential or pragmatic meaning''. In this case a C-unit can comprise "one simple independent finite clause or else an independent finite clause plus one or more dependent finite or nonfinite clauses" (ibid: 310). Freed (1978 cited in Brock 1986:52) notes that a C-unit might consist of a number of "sentence nodes" due to the fact that it may be a run-on or a compound sentence. Alternatively, it might only consist of one word like *yes* when it is an answer to a question like: *Have you ever been sick?* (Loban, 1963 cited in Brock, ibid: 52). Thus, Brock (ibid) prefers the use of C-units for the study of spoken language because they can account for ellipsis. A comparison between the two explanations of the components of C-unit (i.e. the one provided by Foster and Skehan (1996) and the one discussed in Brock (1986)) will reveal the point that they are different from each other because the first one does not consider cases like run-on sentences, coordination, and answers to questions, which might consist of a single word like *yes*.

Criticizing the inconsistent definitions provided for T-units and C-units by different researchers, Foster *et al*. (2000, see also R. Ellis and Barkhuizen, 2005 for a discussion on the topic) have introduced another unit for analyzing speech, namely the AS-unit (the analysis of speech unit). They (ibid:365) defined the AS-unit as "a single speaker's utterance consisting of *an independent clause* or *sub clausal unit,* together with *any subordinate clause(s)* associated with either". Although this definition is based on T-units, it differs from it as it can account for sub-clausal units (Czwenar, 2014) such as *thank you very much* and other speech

phenomena such as false starts, repetitions and self-corrections, topicalizations and interruptions, and scaffolding. This is the reason why this unit is more applicable to the analysis of spoken language.

Having considered all these units, their definitions, and examples of their identification, it does seem to me that researchers have either referred to speech or writing, and that there is a lack of a unit that can deal with the kind of writing that has most of the characteristics of speech such as fragments and sub clausal units as is the case with the data of the present study. Does this type of data need to be treated as speech or writing? Based on this discussion, I argue that it is the data that decides which unit to choose, how to define it, and how to specify whether or not it is error-free (see also Czwenar 2014). For example, neither the literature about the types of units nor that about the studies which have utilized these units can account for cases like these in the corpus of the present study:

e.g. 5.12 S(3) *The saddest even in my life! My name is Araz.*
e.g. 5.13 S(11) *The funniest or saddest it's something that occur in all human being life.*

This noun phrase *The saddest even[t] in my life* cannot be attached to what comes after it as T-unit, *My name is Araz*, because syntactically and semantically speaking it is incompatible with it. As for the second case, example 5.13 could be a T-unit that was easy to identify if the adjectives were not in their superlative form and it would be a correct T-unit: *funny or sad, it's something that occur in all human being life.* However, this would ignore the fact that they are in the superlative form. In addition, it is unlikely that the structure *adjective or adjective+ clause* would be known by a student who is at this level of English. So, even if the adjectives were not in their superlative forms, it would be very unusual for such a structure to occur in the English of a low level student.

Although, as said earlier, the provision of criteria for dividing the text into units and identifying the error-free units from these baseline units depends on the type of data the researcher is dealing with, for the most part I have used Polio's (1997)[3] criteria for the baseline line units and error-free units. However, I have modified them and added more criteria to them to make them more suitable for the data of my research. This is because I found them very detailed and flexible i.e. one could add some and remove some others as deemed appropriate.

5.3 CoA in a number of previous studies

As referred to earlier, researchers first used CoA as a measure of accuracy in language, not as an independent variable but as a dependent one. They have measured correctness to see how it changes as a result of task planning, on-line planning, corrective feedback, task type, etc. For example, Rahimpour and Hosseini, (2010) used the ratio of error-free T-units/total number of T-units to measure the effect of pre-task planning on correctness, and Yuan and Ellis (2003) used the ratio of error-free clauses to the total number of clauses for measuring the effect of on-line planning on correctness (accuracy) as well as complexity and fluency.

Wolfe-Quintero *et al.* (1998) is a very comprehensive survey of the studies that have examined the three dimensions of language development: complexity, accuracy and fluency. They have listed the measures that have been used in different studies for measuring accuracy in language. The following table shows the studies carried out before 1996 together with the accuracy measures they used.

Table 5.1 the studies conducted before 1996, which have used different measures for accuracy in writing (Wolfe-Quintero *et al.* 1998) (including a number of measures for CoA as named in this chapter)

Measure	**Abbreviation**	**Studies**
Error-free T-unit ratio	EFT/T	Arnaud (1992); Larsen-Freeman, (1978); Scott and Tucker (1974) and others
Error-free T-units per sentence	(EFT/S)	Ho-Peng (1983)
Error-free T-units per word	(EFT/W)	Nihalani (1981)
Error-free sentence ratio	(EFS/S)	Tapia (1993)
Error-free clause ratio	(EFC/C)	Ishikawa (1995) and Tapia (1993)
Error-free clauses per sentence	(EFC/S)	Ishikawa (1995)
Error-free clauses per T-unit	(EFC/T)	Ishikawa (1995)
Word in error-free clauses ratio	(WEFC/WC)	Ishikawa (1995)
Errors per T-unit	(E/T)	Flahive and Snow (1980)
First degree errors per T-unit	(1DE/T)	Homburg (1984)
Second degree errors per T-unit	(2DE/T)	Homburg (1984)
Third degree errors per T-unit	(3DE/T)	Homburg (1984)
Errors per clause	(E/C)	Fischer (1984)
Syntactic errors per clause	(SynE/C)	Bardovi-Harlig and Bofman (1989)
Morphological errors per clause	(MorE/C)	Bardovi-Harlig and Bofman (1989)
Lexical errors per clause	(LexE/C)	Bardovi-Harlig and Bofman (1989)
Verb lexical errors per verb	(VLexE/V)	Harley and King (1989)
Lexical errors per lexical word	(LexE/LW)	Engber (1995)

Errors per word	(E/W)	Hyltenstam (1992)
Grammatical errors per word	(GrE/W)	Arthur (1979)
Semantic errors per error	(SemE/E)	Arthur(1979)
Correct connectors per word	(CorrCN/W)	Evola *et al.* (1980)
Correct pronouns per word	(CorrPN/W)	Evola *et al*. (1980)
Correct article ratio	(CorrART/CX)	Evola *et al*. (1980)
Correct definite article ratio	(CorrDef/CX)	Cumming and Mellow (1996)
Correct indefinite article ratio	(CorrINDEF/CX)	Cumming and Mellow (1996)
Correct plural ratio	(CorrPL/CX)	Cumming and Mellow (1996)

A close examination of this table will reveal a number of points: (1) the studies used various measures and either took account of a whole unit such as T-units and clauses, or only a single sentence constituent or structure like articles, pronouns, connectors, and plurals, (2) the studies measured accuracy through either measuring correctness or erroneousness i.e. ratios of error-free units or sentence constituents and structures or ratios of errors whether per word, T-unit, clause, or only per a part of speech like verbs, (3) specific error type counts like grammatical and semantic errors have been used as a measure, (4) error gravity is taken into consideration while measuring accuracy, and this is clear from the method of categorizing the errors by Homburg (1984) into first, second, and third degree errors, and (5) various measures are used in the same study by the same researcher.

Polio (1997) and Polio and Shea (2014) are two other comprehensive reviews of the studies that used different measures of accuracy in writing. Since Polio and Shea (ibid) present an updated list of these research studies, the studies reviewed by them together with their accuracy measures are presented in table 5.2

Table 5.2 the studies that were carried out between 2000–2014 using different measures for accuracy (Polio and Shea, 2014).

Measure	**Studies**
Holistic	Ruegg *et al.* (2011), Evans *et al*. (2010), Barkaoui (2010), Lundstrom and Baker (2009), Storch (2009), Lo and Hyland (2007), Stevenson *et al*. (2006), Ojima (2006), Macaro and Masterman (2006), Lee (2006), McCarthy *et al.* (2005), Evans and Fisher (2005), Sasaki (2000).
Error-free T-units/total T-units	Evans *et al*. (2010), Arslanyilmaz and Pedersen (2010), Storch (2009), Larsen-Freeman (2006), Way *et al*. (2000).
Error-free clauses/total clauses	Kormos (2011), Evans *et al*. (2011), Evans *et al*. (2010), Storch (2009), Ojima (2006), Storch (2005).
Error-free clauses (morphological, syntactic, and lexical errors considered).	Ellis and Yuan (2004)

Percentage of correct noun and verb phrases.	Macaro and Masterman (2006)
Number of lexical, grammatical, and mechanical errors over total words	Ashwell (2000)
Number of errors per 100 words.	Chandler (2003)
Number of errors per T-unit	Kuiken and Vedder (2008)
Errors per words	Storch (2009), Truscott and Hsu (2008), Storch (2005).
Number of specific error types (e.g. articles, prepositions, verb forms, subject verb agreement, grammatical and spelling errors, verb - noun collocation etc.)	Ruegg *et al*. (2011), Laufer and Waldman (2011), Chan (2010), Spoelman and Verspoor (2010), Sheen *et al*. (2009), Bitchener and Knoch (2008, 2010), Bitchener (2008), Bitchener *et al*. (2005) Ellis and Yuan (2004), Fazio (2001).
Error gravity (severity) measures	Evans *et al*. (2014) that dealt with weighted error-free clauses, Kuiken and Vedder (2008) that dealt with T-units with errors in three degrees of severity/total T-units.

This table shows that almost the same measures have been used in measuring accuracy in writing; again some included measuring the erroneous facet of language performance and the others measured the correct facet. Measures considering error severity have also been used. However, a point that is worth taking into account in these two tables is that none of the researchers has combined the same measures in only one single measure, a method that will be tried in the present research study.

It is also important to highlight that one of the problems with the ratio measures of error-free units is that they imply an absolute dichotomy: a unit is either completely right or completely wrong. There is no consideration of intermediate situations. In the case of low-level students, as Polio (1997) argues, an essay may consist only of incorrect units, scoring zero. Moreover, in cases of comparison between high and low level students, these measures may not be so efficient because a good student might make many errors but still produce better language in terms of complexity and even accuracy. For example, two students scoring the same score might be different in the quality of writing they produced because the types of errors that rendered the units of one of them incorrect might be less severe than the types of errors that rendered the units of the other one incorrect. In other words 'incorrect' in both these cases might mean two different things. Consider the following examples:

e.g. 5.14 S(20) *The event was very successful and due to its popularity at that time.*
e.g. 5.15 S(49) *But village people don't similar we.*

Both these T-units have received the same label as 'incorrect' while it is quite clear that the language of subject (20) is at a quite higher level than the language of the subject (49). This is because just removing the *and* from the T-unit in example 5.14 will result in the production of a well-formed T-unit, in a case where this *and* might only be a simple 'slip of the pen' mistake. But the language of subject (49) is more problematic, and it is not possible to correct it with a single change.

Taking these problems into account, researchers have recently begun to consider error gravity in their research (see Evans *et al.* 2014, Polio and Shea, 2014, and Kuiken and Vedder, 2008). This is a positive step even though this method has not achieved high reliability (Polio and Shea 2014). This is why devising a measure that combines a number of these methods might work better in accounting for both high and low levels of students, and distinguishing between the two. This will be attempted in this study.

5.4 The application of the method to the data of the present study

The application of this method involves dividing the text into units. Accordingly, the analysts have to first decide about the measures they are planning to use. For example, the analyst might be interested in the ratio of error-free clauses or T-units; in this case the text has to be divided into clauses, and T-units and each clause and T-unit has to be identified as either erroneous or error-free (see Wolfe-Quintero *et al.* 1998 for a detailed list of the studies that have used accuracy measures). In addition to that, it would be more useful for the sake of consistent application and later replication of the method that the researcher presents a detailed manual of the criteria used in segmenting the text under investigation with real examples from that text. Polio (1997) and Polio and Shea (2014) are good examples of such detailed lists of criteria.

As is clear from section 5.3 there are a large number of measures for assessing accuracy in language. I have decided to use the ratio of error-free T-units to the total number of T-units (EFT/T), the ratio of error-free clauses to the total number of clauses (EFC/C), and the ratio of error-free clauses to the total number of T-units (EFC/T). Hence, I will call this new method from here onward 'T-unit and clause-based correctness analysis'. The reason why I have chosen all the three ratios is because it is not easy to decide at this stage which one could better be applied to the data of my study. Moreover, I have chosen clauses and T-units because it is easier to divide the texts into clauses and T-units rather than sentences due to the huge number of punctuation errors at sentence boundary. However, these punctuation errors

have resulted in many fragments and unfortunately this method takes no account of fragments. That is why, as it will be apparent in the next paragraph, I have used Polio's (1997) criteria that take fragments into account. Another important point is that although the essays that I have collected resemble a spoken style of language, I have not used the units of analysis that are used in the analysis of speech like the C-unit and the AS-unit because the data are meant to be written language and they are supposed to be analyzed as written language. Therefore, it is important to treat the data as they are. Also, I have used T-units to test if a T-unit is a good unit of analysis especially in a method involving dichotomous decisions between correct and incorrect units.

The first step after taking this decision involved dividing every essay into T-units and clauses. Then the error-free T-units and clauses were identified. Breaking down the text into T-units and clauses and identifying the error-free T-units and clauses are based, as mentioned above, on Polio's (1997) criteria (attached as appendix I) in which T-unit is defined as "an independent clause with all its dependent clauses" (ibid:138). These criteria are used because they seemed to suit the analysis of the data of the present study into T-units and clauses. This is due to the fact that these criteria are very strict and clear. Also, these criteria include points that deal with fragments and run-on sentences, and the essays collected as data for the present study contain many similar cases of fragments and run-on sentences. However, not all cases could be treated using these criteria. The problems I encountered even after applying these criteria are described in the following section.

5.4.1 Problems identified while applying T-unit and clause-based correctness analysis

Although Polio's criteria for identifying T-units, error-free T-units and clauses seem to be well suited to the analysis of the present study data, unclear and ambiguous cases still occurred. These cases are identified as problems in the application of T-unit and clause-based CoA as follows:

1. Some fragments occurred where it was not easy or even at times impossible to attach them to what precedes or follows them as T-units or clauses. Polio (1997) and Ishikawa (1995) have referred to fragments. Polio (ibid) allocated one criterion to the cases of fragments that could *appropriately* be attached to what precedes or follows them and this might only mean cases where there is a small mechanical error separating that fragment or phrase from the preceding or following T-unit. Ishikawa (1995) preferred considering a sentence as a unit of analysis (in one case of analysis of her data) rather than T-units because of the high number of fragments in her data (see also Wolfe-

Quintero *et al.,* 1998 for a comment on this). Consider the following examples of fragments that are impossible to attach to what precedes or follows them and fragments that are possible to attach either to what precedes or follows them:

e.g. 5.16 S(20)

Now, you wonder what exactly we talk about, to know the reason for me being so excited to talk about this topic. Well, sometimes we talk about nothing serious, and the conversation becomes just an exchange of what we have done that day and how we feel.

e.g. 5.17 S(10)

I don't like zoo, but when I saw that one I liked it so much and felt comfortable there. Finally, when we came back to the hotel to tidy up our staff, for coming back home. We came to the hotel early in the afternoon, and we had a dinner in the hotel.

e.g. 5.18 S(3) *the saddest event in my life! My name is Araz*.

These fragments are clauses (finite and non-finite) and phrases that are difficult to attach to what comes before or after. The fragment in example 5.16, the non-finite clause *to know the reason for me being so excited to talk about this topic* can neither be attached to the T-unit that precedes it *Now, you wonder what exactly we talk about* nor to the one that follows it *Well, sometimes we talk about nothing serious*. The same case applies to the fragments in examples 5.17 and 5.18. However, notice that the fragments in example 5.19 and 5.20 can easily be attached to the T-units that follow or precede them. Thus, they are attached to what comes before or after them as appropriate and the whole resulting T-unit is considered a T-unit with an error.

e.g. 5.19 S(122)

As I lifted up my foot to kick it. I slipped and knocked to the ground and had one of my ribs broken [the underlined fragment is attached to the T-unit *I slipped and knocked to the ground and had one of my ribs broken]*

e.g. 5.20 S (149)

I will talk about a series of sad events in my life. that make me really upset and for a while I thought I can not go on and continue my live [the underlined fragment is attached to the T-unit before it *I will talk about a series of sad events in my life*]

2. There appeared cases where two coordinated clauses existed in one subordinate clause. The following examples demonstrate such a case:

e.g. 5.21 S(7)

When I started college I faced some difficulties because there were new friends and most of them were spoken in different dilects.

e.g. 5.22 S(22)

The famous country nowadays are USA or UK Because they have many perfect colleges and also they have scientists

It was not easy to decide whether to consider, for example *there were new friends* and *most of them where spoken in different dilects* as one T-unit or two T-units.

3. Some of the T-units are quite correct out of context but they are definitely incorrect in context. Consider the following example:

e.g. 5.23 S(19)

and sometimes your family help you, but you can not tell her.

Out of context, the underlined T-unit could be regarded as error-free but in this case *her* refers to *family*, which is incorrect.

4. Polio has described how to treat cases of punctuation errors at a T-unit boundary but no criterion is allocated to the cases of punctuation errors within sentences. A number of these examples (both punctuation errors within T-units and at T-unit boundary) appeared in the data as in the following examples respectively:

e.g. 5.24 S(5)

student saw another life not like life in school because, life in school was very routin, and easy [punctuation error within T-unit]

e.g. 5.25 S(158)

We never expect what will happen in future, we always plan for something [punctuation error at T-unit boundary]

5. An excessive number of spelling errors (see section 4.3.1 in chapter 4) occurred, which if considered, almost most, if not all the T-units would be wrong and hence none of these measures would work.

5.4.2 More criteria added for the identification of error-free and erroneous T-units

Having considered the above-identified problems in section 5.4.1, the following criteria are added to Polio's (1997) list of criteria:

1. Fragments impossible to attach to what comes before or after them: in the case of the occurrence of a fragment that could not be appropriately attached to any T-unit as it does not seem to be syntactically related to any T-unit, do not consider it for analysis. For example, the fragments *to know the reason for me being so excited to talk about this topic*, *Finally, when we came back to the hotel to tidy up our staff, for coming back home* and *the saddest event in my life!* in examples 5.16, 5.17 and 5.18 above respectively are not included in the analysis while the fragments *As I lifted up my foot to kick it* and *that make me really upset and for a while I thought I can not go on and continue my live* in examples 5.19 and 5.20 above respectively are considered for analysis as they are only separated by a confused punctuation mark and could be attached to what comes before or after them as appropriate. This can be considered a big drawback to using the T-unit as the unit of analysis.
2. Two coordinated clauses in one subordinate clause: when two clauses following a subordinating conjunction are coordinated, consider both of them as a part of the T-unit which contains that subordinate clause. For example, in example 5.21 in section 5.4.1 above, all the sequence *When I started college I faced some difficulties because there were new friends and most of them were spoken in different dilects* is regarded as one T-unit regardless of the fact that there exist two coordinated clauses after *because* (the same applies to example 5.22).
3. Decontextualization: No T-unit should be considered error-free out of context. (see example 5.23 in point 3 section 5.4.1 and the comment about it)
4. Punctuation errors: do not consider any erroneous cases of punctuation within a T-unit as an error (only at a T-unit boundary). To clarify this further, example 5.24 above in point 4 section 5.4.1 is considered one error-free T-unit, although there is an erroneous comma separating *because* from its following clause *life in school was very routin, and easy*, while example 5.25 above in point 4 section 5.4.1 is considered two T-units, one erroneous and the other one error-free; the first one (*We never expect what will happen in future*) is considered erroneous, and the second one (*we always plan for something*) is considered error-free.
5. Spelling errors: do not take spelling errors into account except for those (1) that are morpho-syntactical like the use of *staid* for *stayed* and (2) that lead to the production of

another word not fitting that context e.g. *there* for *their* and vice versa, *weather* for *whether* and vice versa, *were* for *where* and vice versa.

6. Other types of errors: all other types of errors must be considered. As the researcher of this study and an expert speaker of English, I have depended on my knowledge of English to decide whether something is an error or not. However, when I was in doubt, I made use of other sources like discussion with other people who are also specialized in the grammar of English. Phraseology errors were difficult to decide about and hence have been checked in the Oxford dictionary of collocation and British National Corpus (BNC).

So, accuracy or correctness here can be defined as freedom from all types of errors except 'within-sentence' punctuation errors and spelling errors. Also, for a structure to be completely error-free, it has to be correct in context and not out of context.

5.4.3 Dividing texts into clauses

As for breaking down the texts into clauses, every text was divided into three types of clauses:

1. Finite independent clauses (FIC) which can be syntactically defined as consisting of a clear subject and a verb that shows tense e.g. 5.26 *He usually plays tennis with his brother*. When two independent clauses were coordinated by a coordinating conjunction (e.g. and, but, or) they were treated as two finite independent clauses (e.g. 5.27 *they came home and then they had lunch with the family*)

2. Finite dependent clauses (FDC) which consist of (1) a clause marker (e.g. a relative pronoun such as *who*, *which* etc. or a subordinating conjunction such as *because, since, after, before*, etc.) that can sometimes be omitted like the case with relative clause pronouns, (2) a clear subject (apart from cases where the clause marker is the subject of the clause as is the case with the underlined relative clause in example 5.28 and (3) a finite verb that shows tense e.g. 5.28 *the police officer* <u>*who*</u> [clause marker and a subject] <u>*is guarding the house*</u> *is* [finite verb phrase] *famous for his bravery*, e.g. 5.29 *As soon as* [clause marker] *Harry* [subject] *left* [finite verb], *they came in.*

3. Non-finite clauses (NFC) in which there is no clear grammatical subject, the verb is non-finite (does not carry tense e.g. participles, both present and past, and infinitives) and there is another clause element like object: e.g. 5.30 *she went to the village* <u>*to enjoy her time*</u>, e.g. 5.31 *I found* <u>*him working*</u>). The other clause element is added to the definition because

if that element does not exist, the sequence might well be identified as a verb phrase rather than a non-finite clause as in example 5.30 *to enjoy* alone is a verb phrase but *to enjoy her time* is a clause[4].

Another point worth mentioning here is that embedded clauses were first identified within their larger scope (their matrix independent clauses) and then also separately as dependent clauses whether finite or non-finite. However, all complex sentences are directly divided into their constituent clauses (dependent finite and independent finite).

e.g. 5.32 S(62) *[The family tried to encourage me and offered me any thing I wanted only to study [because they Knew that my future is very important for them]*. In this example the sentence is cut up first into the independent finite clause *The family tried to encourage me and offered me any* thing I wanted *only to study* and the dependent finite clause *because they Knew that my future is very important for them* and then into the dependent finite clauses (*I wanted*, *that my future is very important for me*) and the non-finite clause (*to encourage me*).

Three other noteworthy points to highlight are: first, the context was taken into account (like the case with T-units) while evaluating the clauses as either erroneous or error-free; second (unlike the case with T-units) the punctuation confusion and punctuation missing errors at clause boundaries were not considered; and third reduced relative clauses are also considered as FDC. Consider the following examples for these three points:

Point 1 e.g. 5.33 S(3)

I returned like a child who don't know what he is doing what is the true what is the wrong

In this T-unit, the finite dependent clause *what he is doing* is correct out of context but within context it had to be *what he was doing*. Therefore, it was not regarded error-free.

Point 2 e.g. 5.34 S(141)

I myself made an unforgettable trip last week to my home, it was full of surprise, changing life for ever, and an enjoyment.

If error-free T-units were meant for analysis, the underlined T-unit would be rendered incorrect due to the punctuation confusion error represented in placing the comma after the T-unit. But, because punctuation confusion errors at unit boundary are not considered in the case of dividing the text into clauses the underlined clause will be error-free.

Point 3 e.g. 5.35 S(62)

and offered me any thing I wanted only

The reduced relative clause *I wanted* is considered as FDC.

For more clarification of the application of the method of CoA to the data of this study, the following two paragraphs are divided into T-units and clauses and then their error-free units whether T-units or clauses are identified.

e.g. 5.36 S(3)

I woke up and I was in hospital, and I wasn't stable. The whole of my family and her too were beside me. The first thing I asked for "what happened to her". They told that she is fine, but I didn't believe in them. Everyone were beside me, if she fine so who is beside her. After some hours they told me the truth, the hardest truth I had heard. I can't descipe that situation and the pain I got, who will read this essay will understand that pain without description from me. After she died I returned like a child who don't know what he is doing, what is the true what is the wrong. All what I was able to do that was " the life without you has no meaning". I was write these words on walls, books,...
Now after more than one year I am feeling better and step by step I 'm trying to stop thinking of her. Not to forget her but .... It is too hard to descripe. Not to forget her, but not to think of her all the time. After this accident my memory had been bad too. After all what happened to me I didn't lost the life, but I will not stop writing "There life has without you has no meaning".

Table 5.3 the division of the above text into T-units (incorrect and error-free)

T-Units	**Examples**
T	(1) *I woke up* (2) *and I was in hospital* (3) *and I wasn't stable (4) The whole of my family and her too were beside me* (5) *The first thing I asked for "what happened to her"* (6) *They told that she is fine* (7) *but I didn't believe in them* (8) *Everyone were beside me* (9) *if she fine so who is beside her* (10) *After some hours they told me the truth, the hardest truth I had heard* (11) *I can't descipe that situation and the pain I got* (12) *who will read this essay will understand that pain without description from me* (13) *After she died I returned like a child who don't know what he is doing, what is the true what is the wrong* (14) *All what I was able to do that was " the life without you has no meaning"* (15) *I was write these words on walls, books,...* (16) *Now after more than one year I am feeling better* (17) *and step by step I 'm trying to stop thinking of her* (18) *but .... It is too hard to descripe* (19) *After this accident my memory had been bad too* (20) *After all what happened to me I didn't lost the life* (21) *but I will not stop writing "There life has without you has no meaning".*
EFT	1. *I woke up* (1) 2. *and I was in hospital* (2) 3. *and I wasn't stable* (3) 4. *After some hours they told me the truth, the hardest truth I had heard* (10) 5. *I can't descipe that situation and the pain I got* (11)

	6. *Now after more than one year I am feeling better* (16) 7. *and step by step I 'm trying to stop thinking of her* (17) 8. *It is too hard to descripe* (18)

Table 5.4 the division of the above text into clauses (incorrect and error-free)

Clauses	**Examples**
FIC	(1) *I woke up* (2) *I was in hospital* (3) *I wasn't stable* (4) *The whole of my family and her too were beside me* (5) *The first thing I asked for "what happened to her"* (6) *They told that she is fine* (7) *I didn't believe in them* (8) *Everyone were beside me* (9) *so who is beside her* (10) *After some hours they told me the truth, the hardest truth I had heard* (11) *I can't descipe that situation and the pain I got* (12) *who will read this essay will understand that pain without description from me* (13) *I returned like a child who don't know what he is doing, what is the true what is the wrong* (14) *All what I was able to do that was " the life without you has no meaning"* (15) *I was write these words on walls, books,…* (16) *Now after more than one year I am feeling better* (17) *step by step I 'm trying to stop thinking of her* (18) *It is too hard to descripe* (19) *After this accident my memory had been bad too* (20) *I didn't lost the life* (21) *I will not stop writing "There life has without you has no meaning".*
FDC	(1) *I asked for* (2) *that she is fine* (3) *if she fine* (4) *I had heard* (5) *I got (6) who will read this essay* (7) *After she died* (8) *who don't know what he is doing* (9) *what he is doing* (10) *what is the true* (11) *what is the wrong* (12) *what I was able to do* (13) *that …" the life without you has no meaning"* (14) *what happened to me*
NFC	(1) *to stop, (2) thinking of her* (3) *to forget her* (4) *to descripe* (5) *to forget her* (6) *to think of her* (7) *writing"There life has without you has no meaning".*
EFC	1. *I woke up* (1) 2. *I was in hospital* (2) 3. *I wasn't stable* (3) 4. *After some hours they told me the truth, the hardest truth I had heard* (10) 5. *I can't descipe that situation and the pain I got* (11) 6. *Now after more than one year I am feeling better* (16) 7. *step by step I 'm trying to stop thinking of her* (17) 8. *It is too hard to descripe* (18) 9. *I asked for* (1) 10. *I had heard* (4) 11. *I got* (5) 12. *After she died* (7) 13. *what I was able to do* (12) 14. *what happened to me* (14) 15. *to stop* (1) 16. *thinking of her* (2) 17. *to forget her* (3) 18. *to descripe* (4) 19. *to forget her* (5) 20. *to think of her* (6)

Now if we consider the calculation of ratios for this student, it will be as follows:

1. EFT/T= 8 (number of error-free T-units)÷21 (total number of T-units) = 0.380
2. EFC/C= 20 (number of error-free clauses)÷42 (total number of clauses, FIC, FDC, NFC)= 0.47
3. EFC/T= 20 (number of error-free clauses)÷21 (total number of T-units)= 0.95

As obvious from the kind of writing this student has produced and the scores he/she has gained, the student benefits more when clauses are considered because the division into clauses has yielded more correct units than the division of the text into T-units. The ratio is even higher when dividing the error-free clauses by the total number of T-units. This measure is significant because I am looking for a measure that accurately assesses the quality of the student's work. This is necessarily a subjective exercise but, in my opinion, this text can be a good example of an effective piece of writing, which conveys the writer's feelings and emotions. Also, it is fairer if the accuracy of a piece of writing like this can be given more credit through considering smaller units or through describing it in terms of a continuum scale rather than a dichotomous scale.

5.5 User engagement

Applying the method of user engagement (see chapter three section 3.6 for an explanation of this method), another rater, who is also specialized in applied linguistics and thus is more probably going to use the research, reanalyzed 20% of the data in discussion with me as the first rater after being given information about the criteria of coding the data in terms of T-units, clauses (with their three types: finite dependent, finite independent and non-finite clauses) and the definition of error i.e. what kind of errors have been taken into consideration while coding the error-free clauses and T-units. The main aim of this secondary rating was two-fold. The first one was to reduce the level of subjectivity through reflecting on my own coding and changing it where necessary to make it more accurate, and the second was to involve my colleagues in the research that I was doing and I did this through the process of discussion with a fellow teacher. We did not leave any case without reaching a joint decision. A lot of discussion and debate took place between us. The points of discussion included the following:

1. A few cases that were missed in the first time of rating. For example, e.g. 5.37 *I went to my family and I left my sister alone there* produced by S(1) was first coded by me as one T-unit but the second rater realized that it is two T-units.
2. Cases on which we disagreed. For example, the following T-unit was produced by S(5): e.g. 5.38 *I cried for my family* and I coded it as incorrect but the second rater labeled it as correct, and the example 5.39 that was produced by S(8): *While in the college we face a new style of studying which is more simple than the college style* which I coded as correct and the second rater coded as incorrect. The second rater based her decision on the point that *face* and *style* do not collocate.
3. A number of ambiguous cases. For instance, S(10) produced (e.g. 5.40) *and I learnt that they were very different from us in many aspects of life.* I have labeled this incorrect because the pronoun *they* had no clear anaphoric reference in the text. If the reference was to the word *people*, the sentence would be correct but if the reference was to different things like *people, culture and society* as the sentence preceding it was *they have different culture, society, people, and even different thinks*, the sentence will not be correct because it has to be *and I learnt that they were very different from [ours] in many aspects of life.*

Although this is not strictly speaking an inter-rater reliability test, it is interesting nonetheless, to give an account of the degree of agreement between both of us (the first and second rater). The correction coefficients were as follows 0.997, 0.998, 0.999, 0.999, 0.997, 0.998 for T, EFT, FDC, FIC, NFC, EFC respectively.

5.6 Results and discussion

80 essays of the third and fourth year students of two schools (school of Arts and school of Basic Education, 20 students per each level per school) have been first divided into T-units and three types of clauses (FIC, FDC, and NFC), and then these T-units and clauses have been identified as either erroneous or error-free. The next step involved calculating the ratios of EFT/T, EFC/C, and EFC/T for every individual essay of the 80 essays selected (see figures 5.1 to 5.8 below and the table attached as appendix J). All narrative essays (not argumentative) have been considered for this sample of the data. This is to minimize the effect of factors like text type on the performance of the students because it has been found that the type of writing has an effect on the three aspects of language: fluency, accuracy, and complexity. Rezazadeh *et al.* (2011) is a good example of research identifying this. Their subjects performed differently on two different essay writing tasks as far as different aspects

of language were concerned (instruction essay-task group of subjects outperformed the argumentative essay-task group as far as fluency and accuracy were concerned but the effect was reversed in terms of complexity because the participants who wrote argumentative essays produced more complex structures). Table 5.5 shows the range of the three measures in the present study.

As is evident from table 5.5, the lowest results are for the EFT/T. This is most probably because of the strict criteria for error-free T-units (mostly due to the punctuation missing and confusion errors at T-units boundary, which was a very common error). Since this error was not considered for clauses, this measure led to higher averages, though some students still score 0.00. Likewise, Wolfe-Quintero *et al.* (1998) found that there was a huge difference between the means of EFT/T across different studies (0.133 to 0.852), attributing this to the different types of errors considered in each study. I think that the figures in table 5.5 would decrease markedly if spelling errors were also counted because, as the previous chapter shows, spelling errors scored the highest in a sample of 48 subjects. Considering the lowest ranges of students' scores, it can be noted that they are very low; in the case of T-units, they all start from zero. This may show that this method of CoA with dichotomous decisions might not be so effective and useful because it gives less credit to the learners' achievement which might be better shown in smaller units and by continuum assessment especially with the beginners and low level learners.

Table 5.5 the range of all the three ratio measures used

School level	Ratios (lowest to highest)			
	EFT/T	EFC/C	EFC/T	Average of the 3 measures
3rd year students/School of Arts	0.00–0.70	0.06–0.92	0.09–2.26	0.07–1.30
4th year students/School of Arts	0.00–0.53	0.06–0.55	0.08–1.06	0.07–0.66
3rd year students/School of Basic education	0.00–0.68	0.00–0.76	0.00–1.18	0.00–0.88
4th year students/School of Basic education	0.00–0.68	0.095–0.71	0.16–1.04	0.11–0.80

Table 5.5 shows the lowest and the highest measures that the students have scored on the three measures EFT/T, EFC/C, EFC/T and the averages of the three measures.

Figure 5.1 the results of the three measures: EFT/T, EFC/C, EFC/T for third year students/school of Arts: no=20, A= Arts, N=narrative

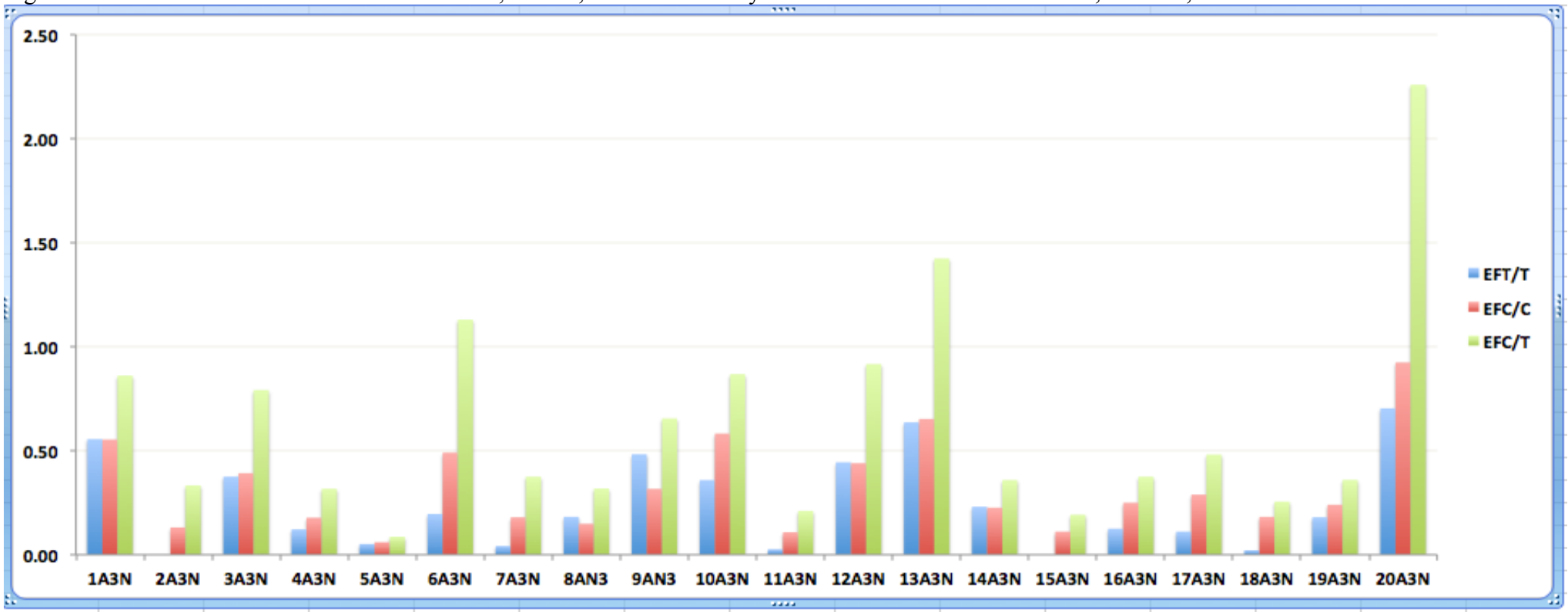
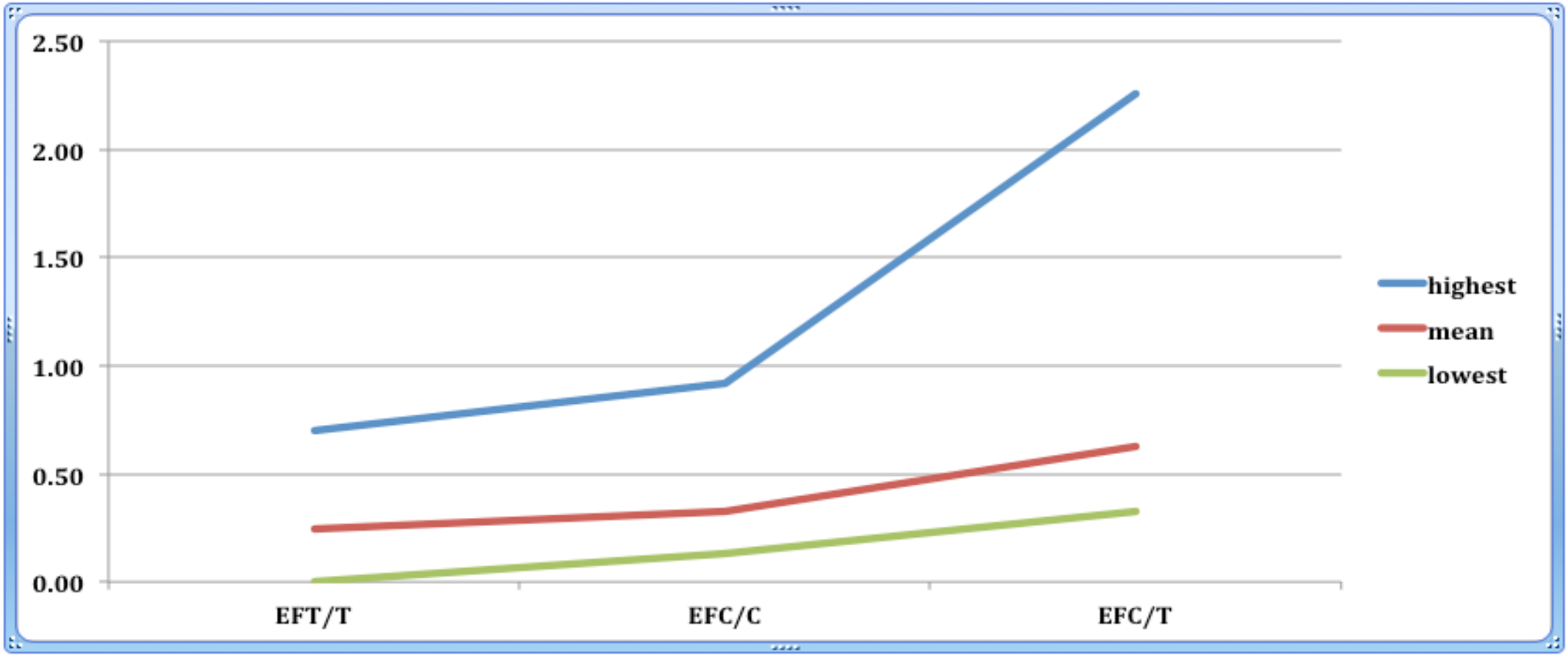


Figure 5.1 shows the results of the three measures: EFT/T, EFC/C, EFC/T for 20 third year students/school of Arts. The horizontal axis shows the codes of the students. For example 1A3N= subject number 1 from school of Arts (A), third year student (3) who wrote a narrative essay (N). The vertical axis shows the ratios (EFT/T, EFC/C, EFC/T) that the students have scored.

In this figure, the results of the three measures vary to a great extent with the EFT/T scoring the lowest and EFC/T the highest. Also, there seems to be a lot of variation among students; they seem to fall into different levels of correctness, and this will become more obvious when calculating the average of the three measures.

Figure 5.2 the results of the three measures: EFT/T, EFC/C, EFC/T for third year students/school of Arts: the highest mark, mean and the lowest



This figure (5.2) is a summary of the figure (5.1) which also shows the mark of the students on the three measures: It incorporates the student who has scored the highest of the students in figure 5.1 and the one who has scored the lowest and the mean of all the twenty scores.

This figure as well shows that the students have scored the highest on the measure EFC/T.

Figure 5.3 the results of the three measures: EFT/T, EFC/C, EFC/T for fourth year students/school of Arts: no=20, A= Arts, N=narrative

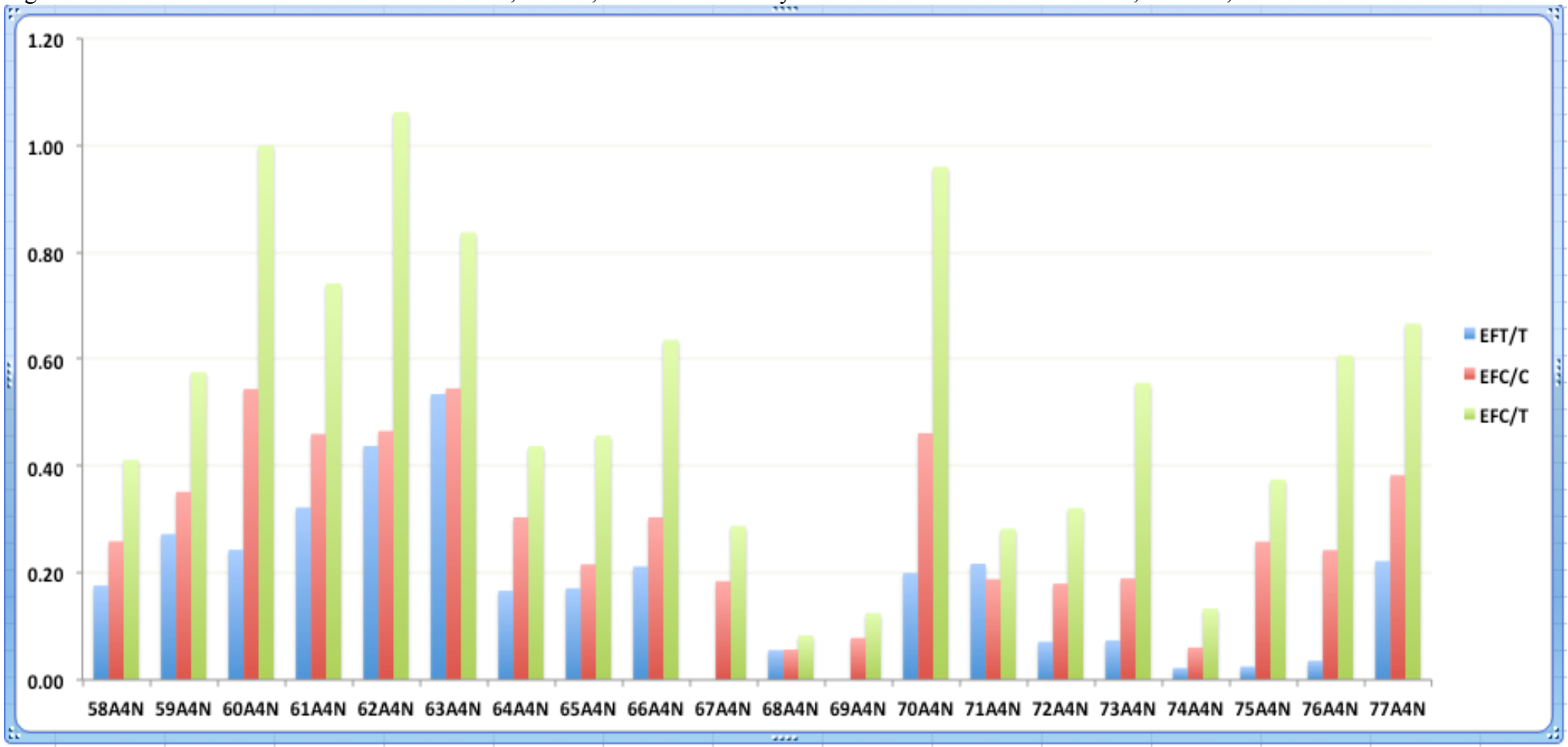
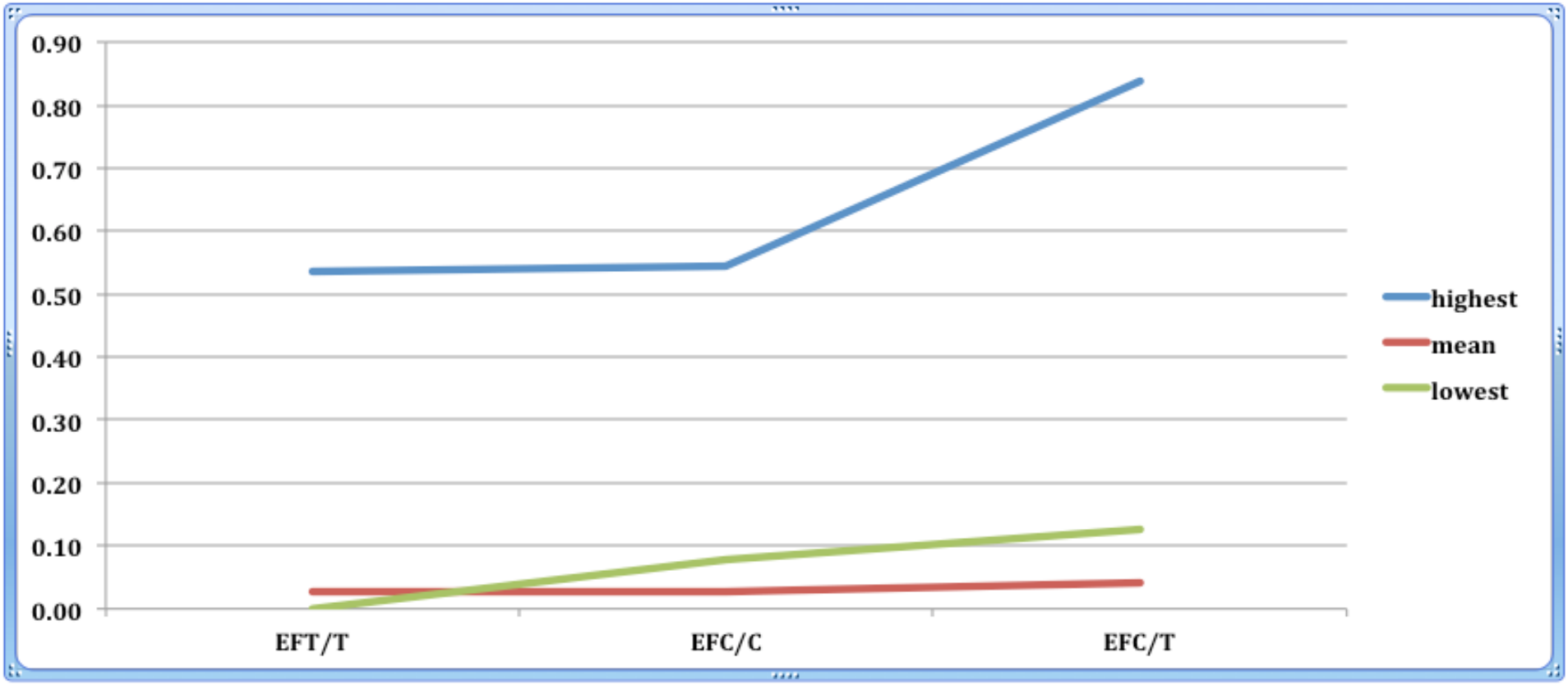


Figure 5.3 shows the results of the three measures: EFT/T, EFC/C, EFC/T for 20 fourth year students/school of Arts. The horizontal axis shows the codes of the students and the vertical axis shows the scores. (For more information about the abbreviations, see figure 5.1).

Again in this figure too, the measure EFC/T scored the highest, and it is quite obvious that there is less variation among the levels of these students.

Figure 5.4 the results of the three measures: EFT/T, EFC/C, EFC/T for fourth year students/school of Arts: the highest mark, mean and the lowest



This figure (5.4) is a summary of the figure (5.3) which also shows the mark of the students on the three measures: It incorporates the student who has scored the highest of the students in figure 5.3 and the one who has scored the lowest and the mean of all the twenty scores.

This figure as well shows that the students have scored the highest on the measure EFC/T.

Figure 5.5 the results of the three measures: EFT/T, EFC/C, EFC/T for third year students/school of Basic Education: no=20, B= Basic, N=narrative

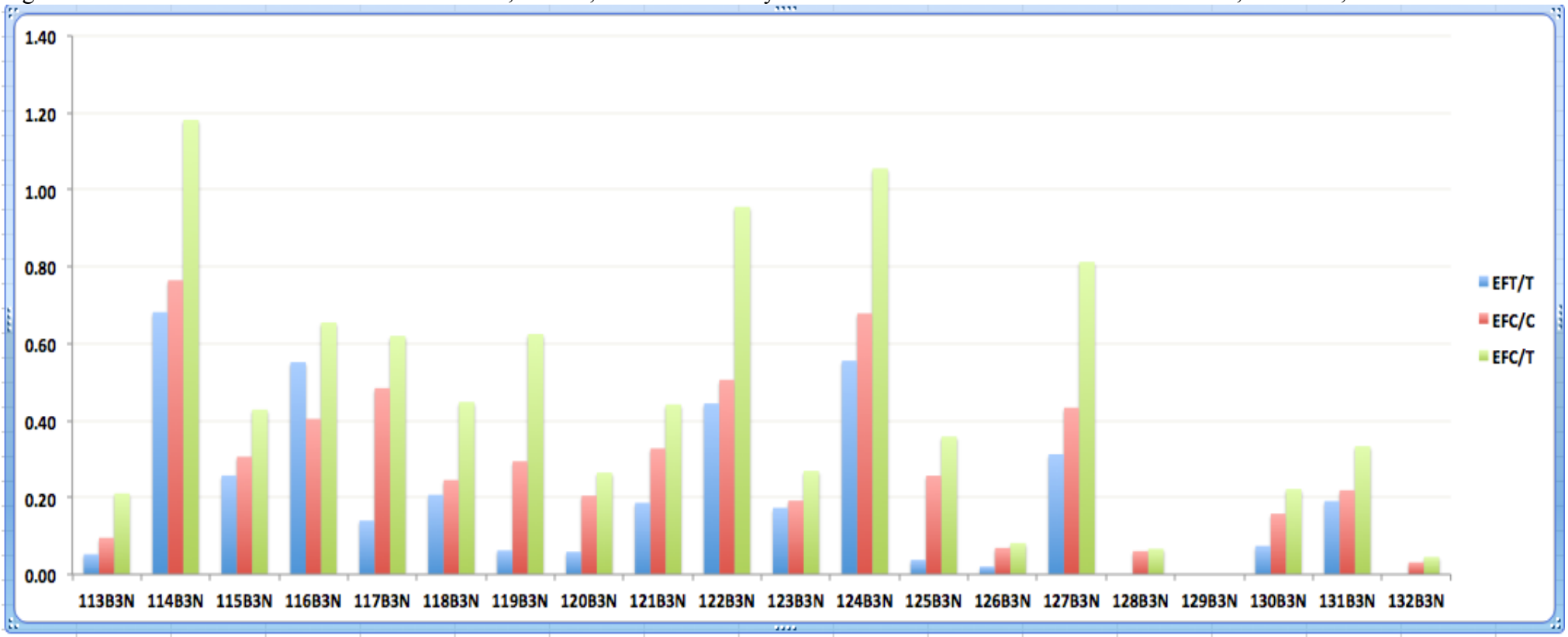
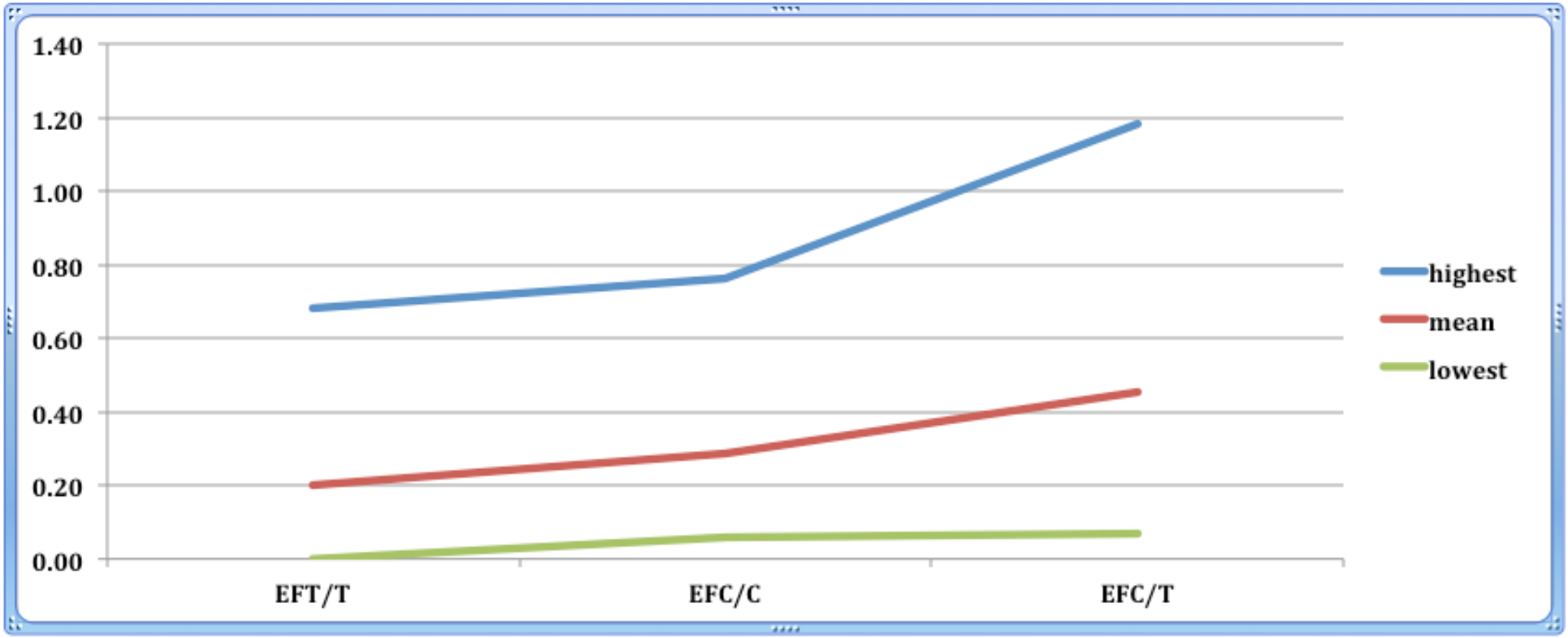


Figure 5.5 shows the results of the three measures: EFT/T, EFC/C, EFC/T for 20 third year students/school of Basic Education. The horizontal axis shows the codes of the students. For example 113B3N= subject number 113 from school of Basic Education (B), third year student (3) who wrote a narrative essay (N). The vertical axis shows the ratios (EFT/T, EFC/C, EFC/T) that the students have scored.

As with the previous figures (5.1, 5.2.5.3.5.4), this figure also demonstrates that the measure EFC/T can better show the students' achievement followed by EFC/C.

Figure 5.6 the results of the three measures: EFT/T, EFC/C, EFC/T for third year students/school of Basic Education: the highest mark, mean and the lowest



This figure (5.6) is a summary of the figure (5.5) which also shows the mark of the students on the three measures: It incorporates the student who has scored the highest of the students in figure 5.5 and the one who has scored the lowest and the mean of all the twenty scores.

Figure 5.7 the results of the three measures: EFT/T, EFC/C, EFC/T for fourth year students/school of Basic Education: no=20, B= Basic, N=narrative

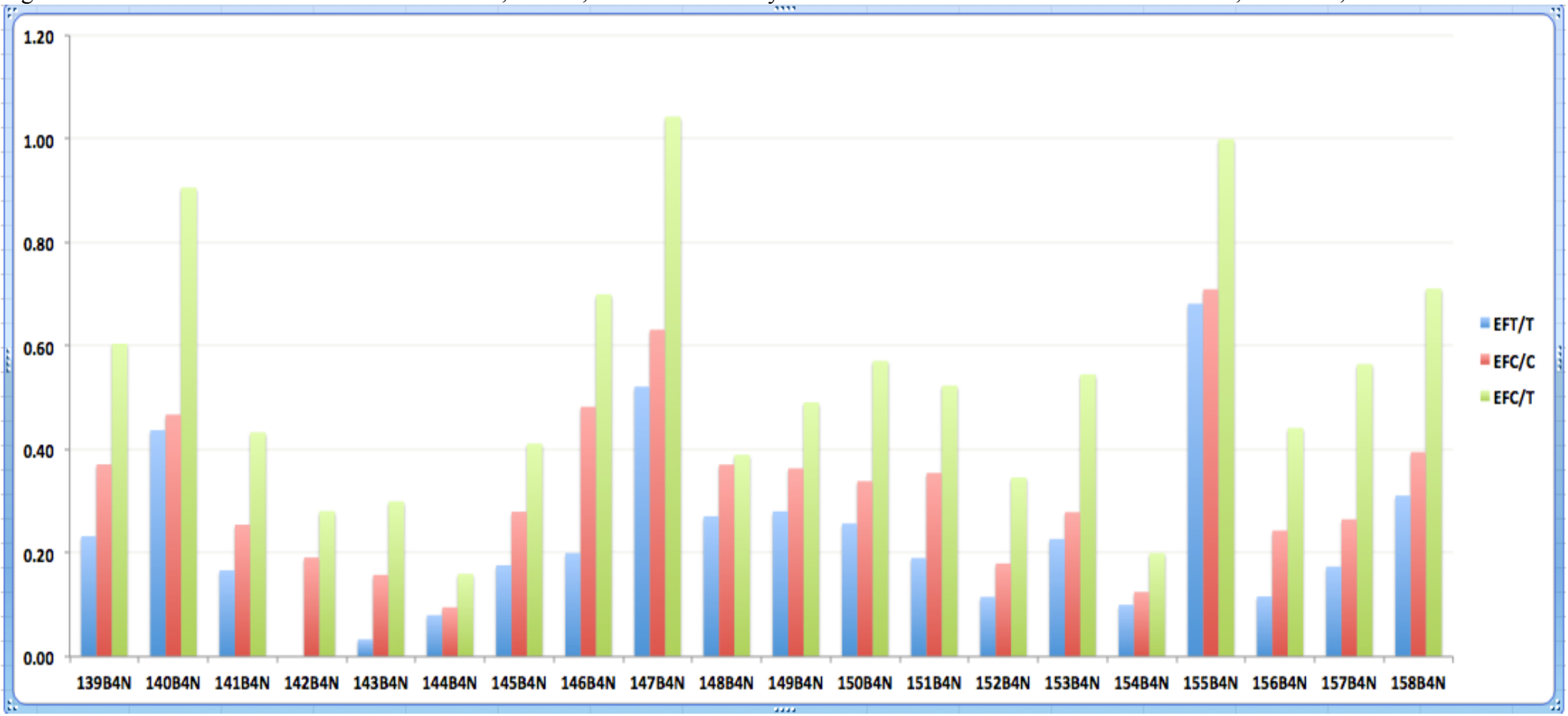
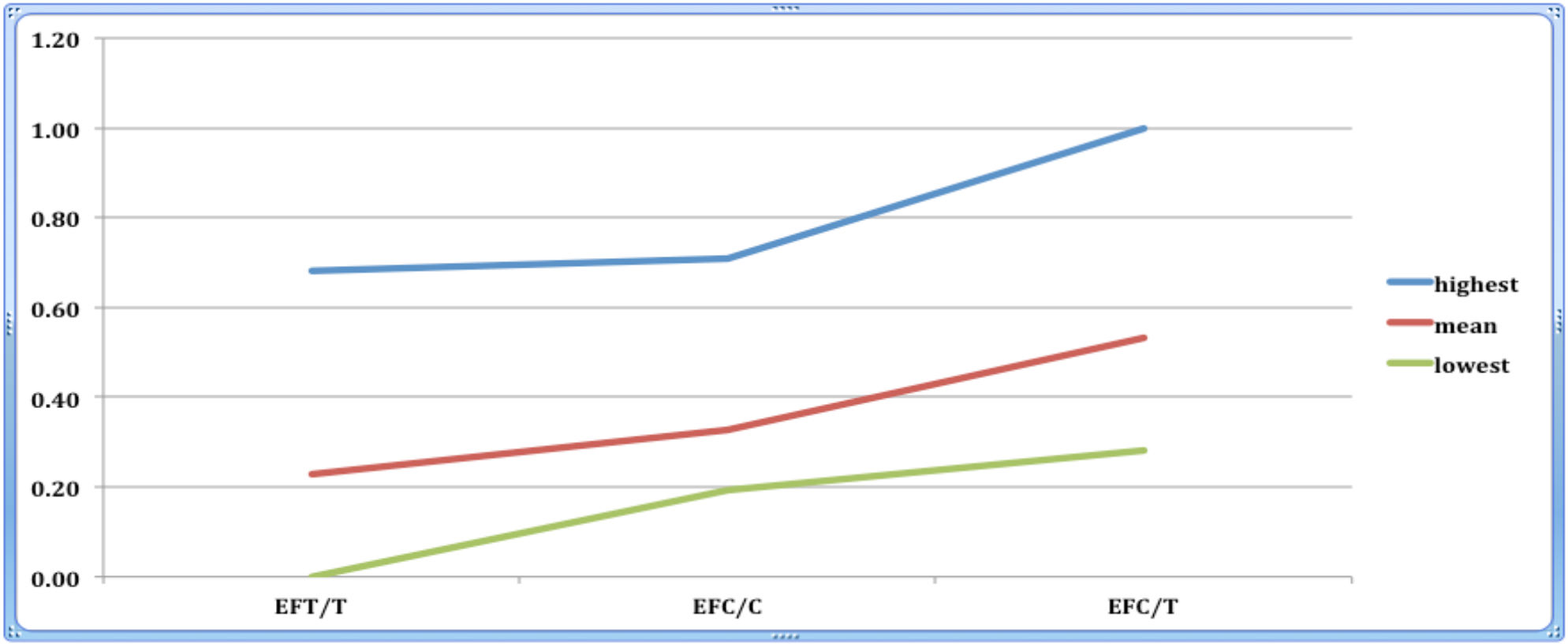


Figure 5.7 shows the results of the three measures: EFT/T, EFC/C, EFC/T for 20 fourth year students/school of Basic Education. The horizontal axis shows the student codes (For more information about the abbreviations, see figure 5.5) and the vertical axis shows the scores.

Like all the previous figures, in this figure the measure EFC/T shows students' achievement more clearly.

Figure 5.8 the results of the three measures: EFT/T, EFC/C, EFC/T for fourth year students/school of Basic Education: the highest mark, mean and the lowest



This figure (5.8) is a summary of the figure (5.7) which also shows the mark of the students on the three measures: It incorporates the student who has scored the highest of the students in figure 5.7 and the one who has scored the lowest and the mean of all the twenty scores.

Like all the previous figures, in these two figures (5.7, 5.8) the measure EFC/T shows students' achievement more clearly

It is very clear from the four figures above (5.1, 5.2, 5.3, 5.4, 5.5,5.6, 5.7 and 5.8) that the three measures are, for the most part, internally consistent. Almost all students do best on the EFC/T, then the EFC/C, then the EFT/T. The students who are very good or very weak are so in all the measures. The reason (besides the strictness in the identification of error-free T-units) why the measure EFC/T can better show the students' achievement followed by the EFC/C (see figures 5.1 to 5.8) is that clauses are "smaller production units for learners, so the domain for achieving error free status is smaller, and scores should be higher than other error free ratios" (Ishikawa, 1995 cited in Wolfe-Quintero *et al*., 1998:47). Moreover, the reason why students have specifically scored the highest on the EFC/T is a mathematical issue. A T-unit is bigger than the clause and that is why there is always a possibility for having fewer of them than the total number of clauses. Additionally, there is always a higher possibility for the students to achieve error-free clauses, hence a bigger chance for their increase. Mathematically, the higher the total of error-free clauses and the lower the total of T-units are, the higher the ratio of EFC/T is. For example, if the total number of error-free clauses is 10 and the total number of T-units is 20 then the ratio of EFC/T would score 10/20=0.5. There may always be a higher possibility for this 10 to increase because students tend to make fewer errors in clauses and a lower possibility for the total number of T-units to increase, as they are big units. Supposing 10 increased to 12 and 20 remained as it is, the total ratio score would be 12/20 = 0.6.

Although these students are supposed to be advanced learners of English, the proficiency level seemed to be very low leading to a high number of fragments that are left unanalyzed as it is the case with Ishikawa's (1995) data. Therefore, T-units are not ideal baseline units for investigating the development in the language of beginner or low-level language learners (ibid). This is also noted by Homburg (1984:89) who mentioned, "compositions receiving lower grades so often have so many errors that the application of objective measures, especially T-unit analysis, is very difficult". However, this might be only true if the researcher is making a dichotomous decision between erroneous and error-free T-units.

I have tried the three measures to indicate that the bigger the unit, the less chance of high achievement by the learner. Indeed, the further one goes with the unit, the greater the probability of incorrectness becomes. The following examples show that the T-unit is incorrect but one of the clauses within the scope of this T-unit is error-free:

e.g. 5.41 S(140)

**Thinking of that the next hour I have a rest, □I spent the whole hour in her office.*

e.g. 5.42 S(127)

**We start talking about our high schools.*

In example 5.41, the whole T-unit *Thinking of that the next hour I have a rest, I spent the whole hour in her office* is not error-free but the clause *I spent the whole hour in her office* is error-free and this is indicated by the asterisk (). Exactly the same occurs with example 5.42; the whole T-unit *We start talking about our high schools* is not correct but the clause *talking about our high school* is correct.

Another issue that may attract attention is that, looking at the figure 5.1 and the table in appendix J, it is quite clear that subjects 13 and 20 have scored the highest on the three measures (Scores for S(13) EFT/T= 0.64, EFC/C= 0.65, EFC/T= 1.42; Scores for S(20) EFT/T= 0.70, EFC/C= 0.92, EFC/T= 2.26). Looking at the background of these two students, some interesting information comes to light. Both subjects have spent some time abroad. Subject 20 was born in the Netherlands and has lived there for quite a long time, and subject 13 has lived in Spain for one year and for 6 weeks in the USA. Most strikingly, she (subject 13) stayed with a family in the USA. As for subject 20, although English is a foreign language for her like the other subjects, she learned her early English in another context, both in terms of her instructed and 'normal daily life' English and this seems to have helped her learn English differently (note that she has scored the highest among all the sample analyzed for this part of the study). Other subjects, also like subject 3 and 1, have declared that they know other foreign languages besides English. For example, subject 3 knows Persian, Turkish, and Arabic and subject 1 knows some German. This might be an indicator that their increased ability to learn foreign languages might have helped them to learn English, which is also a foreign language to them.

Nevertheless, this might not be the case with the subjects that scored the highest in the other school and year four of the same school. There might be something in common between these students which will need further research. A next step, a further research study correlating their average mark in the writing module to what they scored in this task might uncover some useful points in regard to these students.

Figure 5.9 average scores of the three measures: EFT/T, EFC/C, EFC/T for third year students/school of Arts: no=20, A= Arts, N=narrative

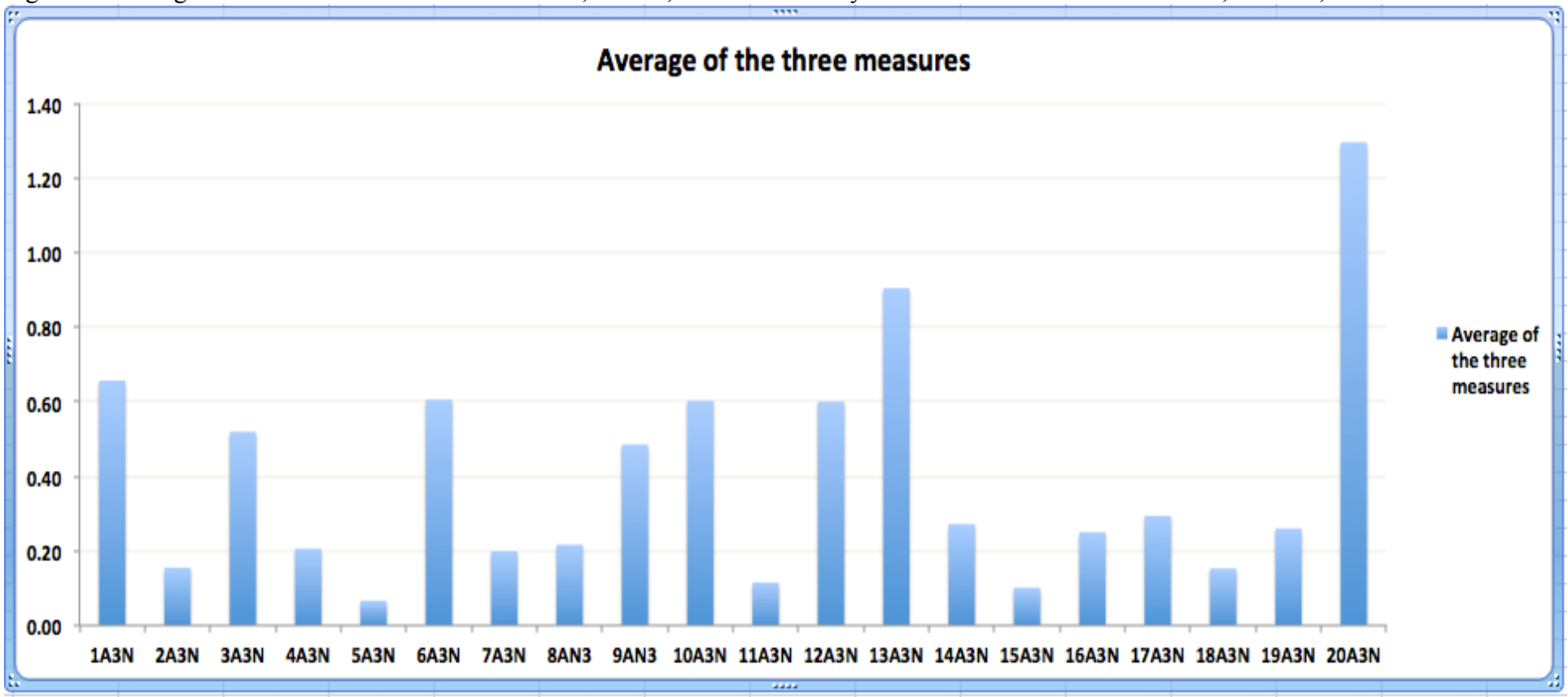


Figure 5.9 shows the average of the three measures: EFT/T, EFC/C, EFC/T for 20 third year students/school of Arts. (For more information about the abbreviations, see figure 5.1).

If we consider only the students who have scored 0.80 and above as advanced, only two students (S20, S13) in figure 5.9 fall in this category. Others can be categorized as intermediate (S1, S6, S10, S12) if we consider intermediate as equal to 0.60–0.79. The rest can be called low-level learners (between 0.00–0.59). It can be noted from the figure that more students fall in the low level category (14 students).

Figure 5.10 average scores of the three measures: EFT/T, EFC/C, EFC/T for fourth year students/school of Arts: no=20, A= Arts, N=narrative

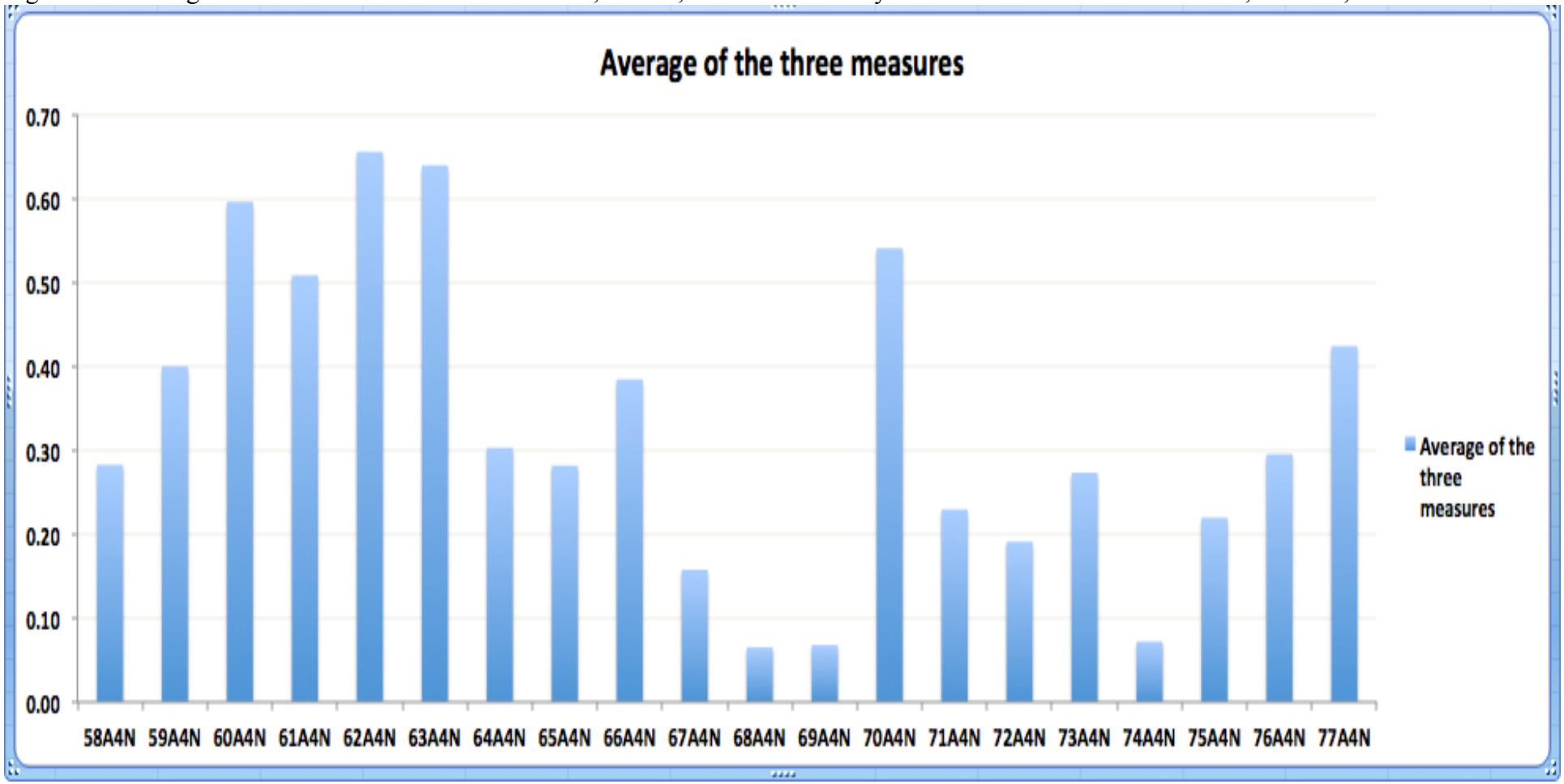


Figure 5.10 shows the averages of the three measures: EFT/T, EFC/C, EFC/T for 20 fourth year students/school of Arts. (For more information about the abbreviations, see figure 5.

If the same ranges of scores are considered for levels, one can see that none of the fourth year students (figure 5.10) can be categorized as advanced and only three students (S60, S62, S63) fall in the category of intermediate. All the other students are low level. To compare the performance of the third year students to that of the fourth year students, although more third year students are labeled advanced and intermediate, the fourth year students in the low level have scored higher than the low level third year students (more students in figure 5.10 have scored close to 20 than in figure 5.9). According to figures 5.2. and 5.4 third year students seem to be better because the highest score in figure 5.2 is higher than the highest score in figure 5.4 and the lowest score is lower. Even the mean scored higher. However, the students in between these numbers in fourth year students are better in achieving higher scores than the third year students. The outliners seem to be in the third year school. Figure 5.10 also shows that this sample of fourth year students seems to be more homogenous than the sample of third year students (figure 5.9) in terms of their level of English. This is quite clear from the smaller variation in the average scores of fourth year students and the considerable variation in the average scores of third year students. This is probably because there are a number of individuals within the third year students who have a good background in English as they have learned their English in a different context or they have a better background in learning foreign languages and this is what makes the third year students seem to have scored higher than the fourth year students even in figures 5.2. and 5.4.

Figure 5.11 Average scores of the three measures: EFT/T, EFC/C, EFC/T for third year students/school of Basic Education: no=20, B= Basic, N=narrative

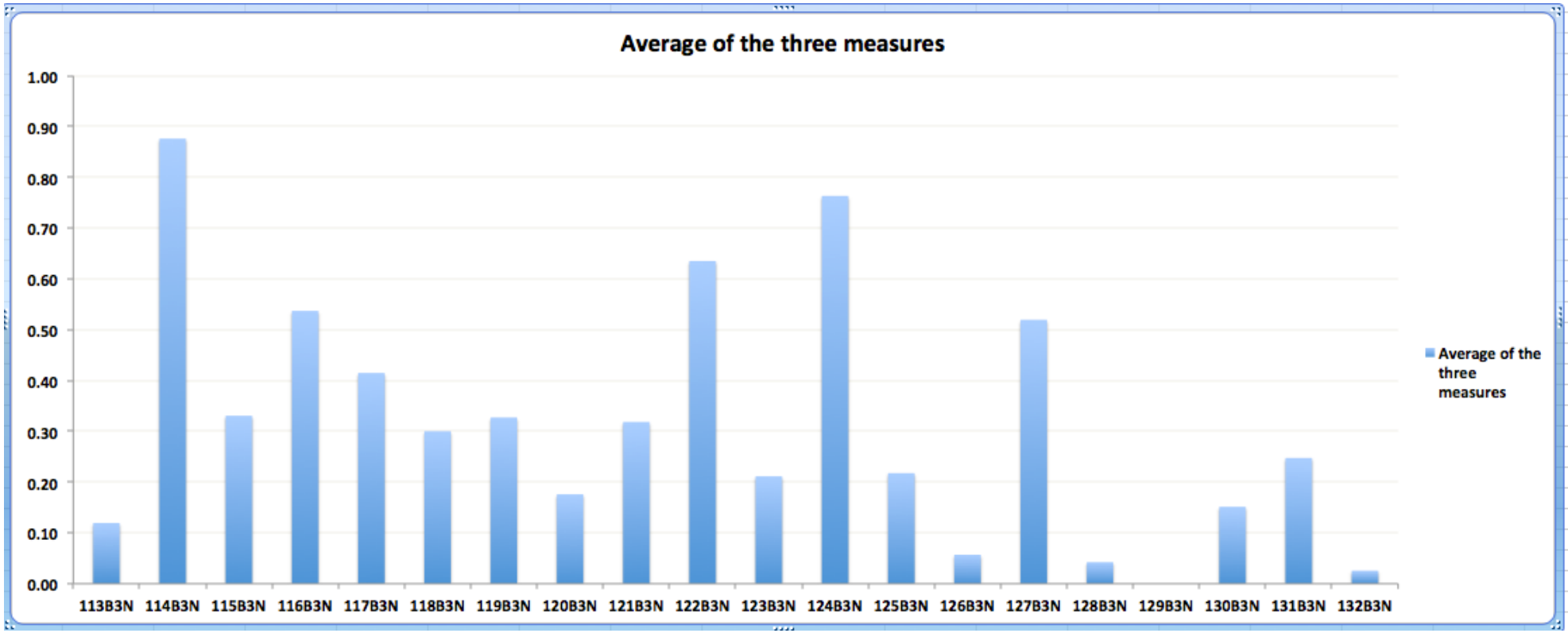


Figure 5.11 shows the averages of the three measures: EFT/T, EFC/C, EFC/T for 20 third year students/school of Basic Education. (For more information about the abbreviations, see figure 5.5).

As for students' performance, only one student (114) can be categorized as advanced and only two can be categorized as intermediate (122,124). All the others have scored below 60 and hence fall in the low level category.

Figure 5.12 Average scores of the three measures: EFT/T, EFC/C, EFC/T for fourth year students/school of Basic Education: no=20, B= Basic, N=narrative

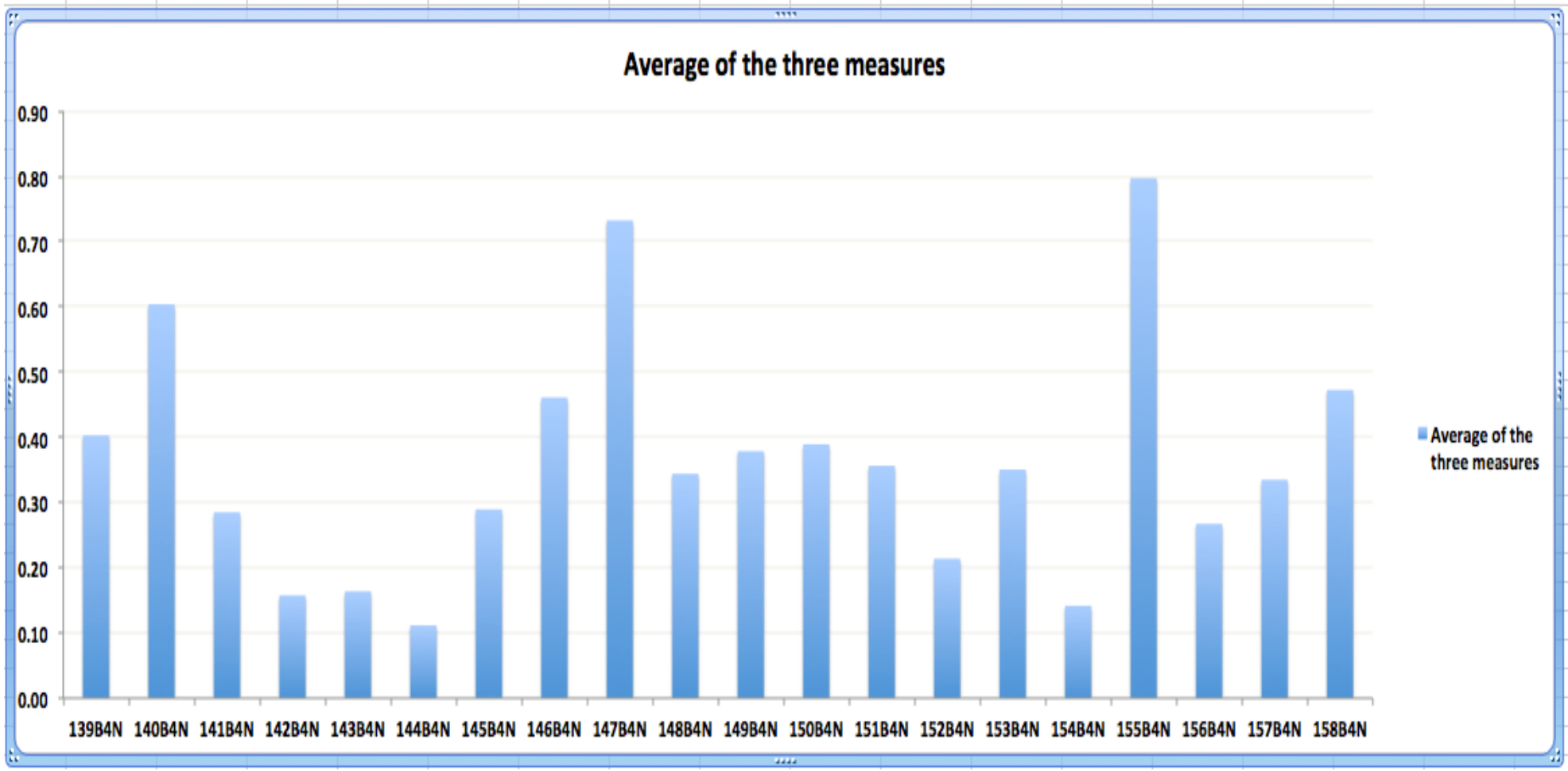


Figure 5.12 shows the averages of the three measures: EFT/T, EFC/C, EFC/T for 20 fourth year students/school of Basic Education. (For more information about the abbreviations, see figure 5.5).

Figure 5.12 shows that only one student (155) has scored the advanced level and only two students (140, 147) fall in the category of intermediate. All the others are low level students. Again with a close inspection of both figures 5.11 and 5.12, it is clear that although the number of advanced and intermediate level students in both groups is the same, overall, the fourth year students seem to have outperformed the third year students because only 4 fourth year students have gained scores below 20 and none of them has scored under 10, while 7 students in the case of third year students have scored below 20 and 4 of them attained very low levels of achievement (4 have scored below 10 and one has scored 0). This is also obvious in figures 5.6 and 5.8. as in figure 5.8 , although there is no outliners, the mean and the lowest score are higher than in figure 5.6.

Overall, the fourth year students tend to be slightly better than the third year students. The case in the third year group of the school of Arts is not very straightforward because there were individuals who have had high scores due to being abroad and this led to higher average scores than the fourth year students (0.39 compared to 0.32). This slight improvement in the performance of the students from the third year to the fourth year has been tested statistically using a 2-tailed T-test and the results were not significant (P=0.35 for school of Arts) (P= 0.48 for school of Basic Education). This indicates that year of study did not have a very significant effect on the performance of the students.

5.7 Conclusion

This chapter has dealt with CoA as a method for measuring accuracy. It has reviewed the units of analysis that those researching CoA have used, the research conducted in this field, and the ratio measures this research has made use of in calculating correctness. It became clear from the review of the literature that several units have been used for measuring CoA in language. The type of units used in the analysis seems to depend on the genre or the type of the text the researcher analyzed, more specifically whether the text is written or spoken. The units analyzed in the written texts included T-units, clauses and sentences while the units analyzed in speech included (in addition to T-units) C-units and AS-units, which account for the speech elements such as sub-clausal sequences, ellipsis, false starts, repetitions etc. Although researchers have treated speech and writing differently as regarding the use of units of analysis, none of the research studies that are reviewed in this chapter mentioned the type of writing that has many characteristics of speech like sub clausal units, fragments, direct

quotations etc. In this case, what will be useful to use as a unit of analysis? The units used in calculating the accuracy in speech or writing or a different unit that can deal with both?

The chapter has also reviewed the literature on the ratio measures used for measuring correctness in writing. There are two vital points to shed light on here. The first one is that researchers who investigated accuracy in writing used a number of various measures including EFT/T, EFC/C, EFS/S, EFC/T etc. The second one is that although a lot of research studies are carried out in the field, few of them have measured the reliability of their methods and also very few have presented clear criteria for identifying both erroneous and error-free units of analysis (see Polio and Shea, 2014; Foster *et al.* 2000 and Polio, 1997).

The practical part of the chapter applied three ratio measures of correctness, namely EFT/T, EFC/C, EFC/T to the data of the present study in order to answer the question raised in the introduction "can a number of these measures be usefully applied to the data of the present study?" It appears that this can be done with the support of clear guidelines and after treating the problems faced while tabulating the text into units and identifying the error-free from these units. One of the measures of usefulness could be distinguishing between the good students and the bad students and describing more objectively the kind of English these students have written. It is evident from the comparison between the performance of the students (the two levels of both schools, school of Arts and school of Basic Education) that the measure EFC/T can more precisely measure accuracy. This is because clauses are smaller units and their domain of errors is small and hence less prone to errors, and the T-units are bigger units, hence more possibility for a lower total number. Thus, when the number of error-free clauses increases and the number of T-units decreases, the ratio EFC/T also rises and students are given more credit for having produced more error-free units. As for taking account of the kind of English the students have written, the measures could not successfully give a clear picture of how well they actually performed because of not considering error gravity.

As for the performance of third year students compared to that of the fourth year students of both schools in terms of how accurate they are in writing, the third year students are less homogeneous in their level of performance than the fourth year students. However, although a number of the third year students outperformed the fourth year students in terms of advanced and intermediate levels, overall, the fourth year students scored higher than the third year students. However, the difference between the two groups is not significant. This means that the year of study does not have a great impact on the performance of the students.

One other noteworthy issue is that a new method of rechecking the quality of the analysis is used in my study, which I have called 'user engagement'. A second rater reanalyzed 20% of the data analyzed for this chapter in a discussion with me as the first. This revealed important points of discussion within which there were points of agreement and disagreement.

One significant point remains to be highlighted. Despite the fact that the measure EFC/T could distinguish between levels of students, with the low level students (like the subjects of this study), smaller units than clauses and T-units might be of higher benefit and most probably will indicate the level of students better than the longer units and will cover more of the students' correct language; this is especially true when one is making a dichotomy between error-free and erroneous units regardless of the severity of error(s) that make a given unit erroneous. This can be ascribed to the fact that the shorter the unit, the less prone it is to errors, especially with the low level students. This is why, in the next chapter, different units will be considered. These units could be phrases, clauses, or sentences.

Notes

1. I have given this acronym to the correctness analysis to differentiate it from the CA which is an abbreviation that stands for 'contrastive analysis' as also opposed to error analysis.
2. Some of these coordination problems have been dealt with in the analysis of the corpus of the present study (see point 2 section 5.4.1).
3. The criteria are attached as appendix I.
4. See Huddleston & Pullumn, 2005; and Quirk & Greenbaum, 1973 for the definition and examples of finite and non-finite clause.

Chapter Six

Various-Units-Based Correctness Analysis

6.1 Introduction

Like the previous chapter, this chapter focuses on measuring correctness rather than erroneousness. It aims to (1) test a new method of analysis and a new measure of correctness, (2) describe that method and measure in detail, in contrast with the method described in chapter 5 and (3) describe the level of performance of the third year students in comparison to the level of performance of the fourth year students, according to this new measure. Before moving to the outline of this chapter, it is important to highlight that I have called the method 'various-units-based correctness analysis' because it is based on different units of analysis.

The chapter begins with a detailed description and exemplification of this new method. It then presents the results of applying this method, this measure and the T-unit measure to two essays as examples in order to illustrate the difference between this method of analysis and measure of accuracy and the methods and measures used in the previous chapter. Because I faced a number of difficulties while applying the method, I have summarized these difficulties in section 6.5. As this method is used in the analysis of 80 essays written by the third and fourth year students of two schools and as a measure for calculating their accuracy, section 6.6 is devoted to the description and discussion of the results that the analysis yielded. The chapter ends with a concluding section.

6.2 Description of the method of various-units-based correctness analysis

In essence, this method analyzes error-free units; it is therefore a method of correctness analysis. It is similar in this regard to the method described in the previous chapter, which also depends on identifying error-free units. Furthermore, the measure used with this method is also similar to two of the measures used in the previous chapter: EFT/T and EFC/C in the sense that its calculation is between 0 and 1. However, there are some points of difference between the method described in the previous chapter and the method described here (the various-units-based correctness analysis). With the T-units and clause-based correctness analysis method (1) the text is first divided into units like T-units and clauses and then the

error-free instances of these units are identified. In other words, the analysis is a top-down analysis, (2) the units the text is divided into are similar in their syntax i.e. they are based on one syntactic definition of their composition. For example, the identification of all T-units is dependent on one clear definition and a set of well-defined criteria. In contrast, with the method I call 'various-units-based correctness analysis' the text is divided into small and/or big correct units that are not similar in their syntax. They might be sentences of different types, clauses, phrases, or multi-word units or the unit might even consist merely of the subject and verb components of a clause, where the object is present but incorrect. One can call this analysis a bottom-up analysis as it starts from any sequence no matter how small. The measure of accuracy that is used in relation to this method involves multiplying the number of the identified correct sequences (units) by their mean length and divided by the total words in the essay resulting into a score between 0 and 1.

$$accuracy = \frac{[no.of\ correct\ sequences\ (CS) \times (mean\ length\ of\ correct\ sequences\ (\ No.of\ words\ in\ CS/\ NCS\ )\ ])}{total\ number\ of\ words\ in\ essay}$$

Or accuracy= CS*MLCS/TWs per essay

The following are some points that explain the application of this new method:

1. The aim is to mark all the correct sequences;

2. A sequence could be taken to mean a unit of any size, or any sequence of words belonging to any unit. In my application of the method, a sequence is a structured unit rather than a series of words. This is because I view language as hierarchically structured and acknowledge that teachers of English have to take account of units. I could simply take any string of words as correct. Consider the following example:

   e.g. 6.1 S(71)
   some good friends who help [=helped] me a lot

   This phrase contains a tense error; in context, 'help' should be 'helped'. How much correct language is there here? If the longest correct string is taken, irrespective of structure, then 'some good friends who' would be the correct sequence. However, I prefer to take only units that are well-structured and hence only the phrase *some good friends* has been considered correct, excluding the word *who*. This is because taking any string of words to be correct (if it does not violate the rules of the target language) would not do justice to what the students have done and would allow too much to be correct.

Therefore, on balance, having considered the structural view of language and the sequential view of language, I came to the decision that I will take the grammatically well-structured sequences (that syntactically make sense) as units of analysis.

3. A correct sequence can be a stretch of a minimum of two words, which are structurally well-formed i.e. sequenced grammatically. It can be complete or incomplete.

3a. A complete sequence can be either (A) a phrase (noun phrase: e.g. 6.2 *the most important point*, e.g. 6.3 *very kind people*; verb phrase: e.g. 6.4 *will admit*, e.g. 6.5 *would have forgotten*; adjective phrase: e.g. 6.6 *very big*, e.g. 6.7 *terribly sorry*; adverb phrase: e.g. 6.8 *very quickly*, e.g. 6.9 *quite hesitantly*; prepositional phrase: e.g. 6.10 *in the room,* e.g. 6.11 *behind the man*); (B) a clause (non-finite: e.g. 6.12 *Playing music*; e.g. 6.13 *having been helped by his father*, finite independent: e.g. 6.14 *he went home*, e.g. 6.15 *they approved the proposal,* or finite dependent e.g. 6.16 *because he arrived late, he missed the train;* or (C) multi-word units e.g. 6.17 *day by day*, e.g. 6.18 *as simple as* etc.

3b. An incomplete sequence has an incorrect element (e.g. the object or the complement); for example, a transitive verb with either an incorrect object or subject, or an intensive verb with an incorrect subject or complement or verbs like *think, say, realize* followed by *that* but still incorrect complement: e.g. 6.19 S(1) *he used to love German and their language* (intended *he used to love Germans and their language*), e.g. 6.20 S(19) *it was one of my dream*, (intended *It was one of my dreams),* e.g. 6.21 S(115) *There are many journey in my life (*Intended *there are many journeys in my life),* e.g. 6.22 S(7) *She said to us that the system of education will be courses,* (intended *She said to us that the system of education would be courses)*[1] e.g. 6.23 S(14) *I realized that I am like all the students*, (intended *I realized that I was like all the students*), e.g. 6.24 S(2) which made me doing things (intended *which made me do things*)

If one wants to identify the structure (the syntactic pattern) of the incomplete sequence, it is possible to write it as a normal pattern but with the wrong item placed between brackets. For example, it can be SV(C) when the complement is wrong or SV(O) when the object is wrong or SVO(V) as in the case of the causative verb (make) in example 6.24: *which made me doing things* etc.

4. All errors, including lexical and phraseological, are considered except for spelling errors and punctuation errors (see example 6.27 below for punctuation errors). For

spelling errors, only morphological errors are taken into account and when the error involves the production of a wrong word, such as *accept* for *except*, *weather* for *whether*, *head* for *had*, *intend* for *attend*). All other errors are considered and if they are found, the sequence is regarded as erroneous.

5. A *correct* sequence means a stretch (any of those exemplified above in 3a) that is completely correct, i.e. without errors except for spelling and punctuation errors as mentioned in point 4 above.

e.g. 6.25

a helpful friend but not * *helpful friend,* e.g. 6.26 *they were in the room* but not **they was in the room.*

The following paragraph (e.g. 6.27) is given as an example to show the process of identifying correct sequences in a given text. (Note that punctuation errors are not taken into consideration). The underlined sequences are correct and those that are not underlined are incorrect.

e.g. 6.27 S(16)

You will meet a new friends some people that you don't know any thing about so at those time I as apersonality faced some problems or some thing that are bitter for me but thing like those will change by the time when I first came to college every man were shy and have unhappy feelings at school there was lessons at college was lectures before college I have never meet a doctor or persons like these at school there was no dean, head of department or staff.

6. As for identifying the boundary of a given sequence, it can go as far as it is correct, bearing in mind the fact that it has to be one of the structures mentioned in 3a.

With the above example (6.26): *they was in the room*, it is only *in the room* that will be included in the analysis as a completely correct sequence. Consider the following example also:

e.g. 6.28 S(2)

But I standed toward of all these problems and I never thought of giving up for my big wish which I have for learning English language, although my mind had a lot of thought which made me doing things late and killed my time .

7. If two sequences are coordinated by a coordinating conjunction, the boundary of a sequence can be extended as far as the whole coordinated structure provided that this coordinated structure is correct and coordinating it out of context makes sense.

e.g. 6.29 S(8)

my idea and point of view

In example 6.29, the two parallel coordinated sequences are taken together and count as one sequence

e.g. 6.30 S(1)

I went to my family and I left my sister alone down there.

In example 6.30 too, the two coordinated structures *I went to my family* and *I left my sister alone down there* are taken together because they are two parallel structures and coordinating them out of context makes sense.

8. In the case of erroneous coordinated sequences, whether sentences or phrases, or when the coordination between the two sequences does not make any sense if taken out of context, they are divided into their constituents and the coordinating conjunction is not included in any of the coordinated sequences as in the following examples:

e.g. 6.31 S(9)

many people try to get good marks and they do hard in their life and there are people who get accellent degrees

Now joining the phrase *in their life* and the clause *there are people who get accellent degrees* does not produce a well-formed sequence. In consequence, this is taken as three separate units.

e.g. 6.32 S(9)

while there are students who see the first everyday and they do not learn new things in their college and those studend see difficulty in their in their life of college, especially those who do not care about knowledge,

In this example, the clause *they do not learn new things in their college* is preceded by the coordinating conjunction *and* but it is not taken with the clause because the sequence before this clause is wrong. Separating coordinated structures in this way increases the number of 'correct' units identified.

9. In case of run-on sentences and fragments, the boundary of the sequence is identified where the confused punctuation mark is placed. In cases where the punctuation is absent, the sequences are separated where the absent punctuation mark should have been placed.

e.g. 6.33: S(5)

First week at college It different for all students

e.g. 6.34 S(7)

The First day at college at the beginging I was always thinking of my friends who were with me at school and time to time I phone them to ask what are they doing. Then I met a student who named Farhad, we introduced ourselves to each other, Farhad was ohnest to me and helped me to know other students.

As evident in examples 6.33 and 6.34, the boundaries of the sequences lie either where the punctuation is identifiably absent, or where it is confused (one punctuation mark is put in the place of another like comma in the place of full stop) like the boundary between the noun phrase *First week at college* and the erroneous finite clause *it different for all students* in example 6.33, between the noun phrase *The First day at college* and the finite clause *at the beginging I was always thinking of my friends who were with me at school and time to time I phone them to ask what are they doing,* and between the three finite clauses *Then I met a student who named Farhad* , *we introduced ourselves to each other* and *Farhad was ohnest to me and helped me to know other students* in example 6.34.

10. Fragments: fragments are also analyzed in this method. Consider this example:

e.g. 6.35 S(73)

Although my other sisters and brothers did not love him as much as they loved my mother, because my father selected me and my brother as his beloved children.

Example 6.35 consists of two subordinate clauses but no main clause. If T-units are considered, then, it would not be an error-free T-unit. However, we can isolate the two clauses without their subordinating conjunctions (i.e. *my other sisters and brothers did not love him as they loved my mother; my father selected me and my brother as his beloved children*) and identify these as correct without implying that the clause combination as a whole is correct. This is another point in favor of this method.

11. Contextualization: Errors are contextualized i.e. the context and co-text are taken into account while accounting for errors.

e.g. 6.36 S(2)
I tried to achieve that step.

Here, in this example, there is a phraseological error: A *step* cannot be *achieved,* so the sequence is divided into *I tried to* and *that step.* Both of these sequences are correct out of the context of the whole clause although within the clause they are not.

Taking a phrase out of context is illustrated also by examples 6.37 and 6.38. In example 6.37 *in Turkey* is an acceptable sequence on its own, even though in context it should be *to Turkey.* In example 6.38, if this sentence is taken out of the context of the whole essay, it is correct. However, in the context of the essay, there is a tense error: the sentence has to be in the past tense. That is why, the sentence is broken down into only two sequences *all students* and *at me*.

e.g. 6.37S(10)
Last year I with my family went to a journey in Turkey.

e.g. 6.38 S(130)
All students are looking at me

12. Word calculation: Numbers are not counted when they are written as digits but all words are counted. Compound words are counted as one if written together and two if written separately.

6.3 Two examples

1. Example. 6.39 S(9): This example is an essay produced by the subject number 9. This essay is attached as appendix L. It is tabulated (table 6.1) according to this method and the method of T- unit analysis to show the difference between the two methods.

Table 6.1 the division of the text in example 6.39 into correct sequences using the various units-based correctness analysis.

	Correct sequences	**Structure of correct sequences**
1	It was a week	Simple sentence
2	I met new friends	Simple sentence
3	That week was	S V(C)
4	because of many new things that I had not seen in the life of school	Finite dependent clause
5	The first week at college is not	S V(C)
6	because there are many things which are unexpected, and sometimes we have in the life of school simple things and simple work	Finite dependent clause/ with finite independent clause
7	While in the college we face a new style of studying which is more simple than the college style	Simple sentence
8	No one knows how many people	SV(O)
9	in their last year of school	Prepositional phrase
10	their first week of college	Noun phrase
11	with happiness	Prepositional phrase
12	the point is	SV(C)
13	the college	Noun phrase
14	of course they will first meet new people	Simple sentence
15	the first week I came to the college was a week full of happiness because I met new friends and new teachers and that week was full of interesting things and happy moments	Compound complex sentence
16	I felt happy day by day	Simple sentence
17	I was afraid of how to go on with the new style of studying, but now I am going on in college and it is simple for me	Compound sentence
18	on one hand the first week is a week of saddness because of the difficult and the different style of having lectures and many other things but on the other hand it is a week of happiness	Compound sentence
19	the college	Noun phrase
20	many other things	Noun phrase
21	There are many reasons behind why the first week is important at college and the most important reason is how to go on in a good way to get high marks	Compound sentence
22	many people try to get good marks	Simple sentence
23	in their life	Prepositional phrase
24	there are people who get accellent	Simple sentence

	degrees	
25	may come from how to be	Verb phrase
26	from the first week	Prepositional phrase
27	how to go on in the following weeks	Dependent finite clause
28	the first week at college is a normal week	Simple sentence
29	for those who are clever	Prepositional phrase
30	in their life	Prepositional phrase
31	they see the firs week as a normal week	Simple sentence
32	more difficult weeks in their life	Noun phrase
33	those who are on the top	Noun phrase
34	There are students	Simple sentence
35	Who see	SV(O)
36	they do not learn new things in their college	Simple sentence
37	in their life of college	Prepositional phrase
38	especially those who do not care about knowledge	Noun phrase
39	it is not as simple as	SV(C)
40	So the first week at college is a week of difficulties	Simple sentence
41	that is it.	Simple sentence

The application of the above-mentioned measure of accuracy to this example (6.39) yields the following results:

Accuracy of essay no. $9 = 41 \times [366 \div 41] \div 453 = 0.80$

Applying the division of the text into error-free and erroneous T-units to example 6.39, as shown in the following table 6.2, however, showed that only 16 T-units are correct from a total number of 30.

Table 6.2 the application of the T-unit-based analysis to the essay in example 6.39

	All T-units	Correct T-units
1.	It was a week which I met new friends	and sometimes we have in the life of school simple things and simple work.
2.	That week was a worried week because of many new things that I had not seen in the life of school	While in the college we face a new style of studying which is more simple than the college style
3.	The first week at college is not like other weeks because there are many things which	the first week I came to the college was a week full of happiness because I met new friends and new teachers

	are unexpected	
4.	and sometimes we have in the life of school simple things and simple work	and that week was full of interesting things and happy moments
5.	While in the college we face a new style of studying which is more simple than the college style	I felt happy day by day
6.	no one knows how many people are there in their last year of school who are waiting to their first week of college with happiness	I was afraid of how to go on with the new style of studying
7.	but the point is how will they enter to the college and what should they do before they enter to it	but now I am going on in college
8.	of course they will first meet new people, new student and teachers	and it is simple for me.
9.	the first week I came to the college was a week full of happiness because I met new friends and new teachers	on one hand the first week is a week of saddness because of the difficult and the different style of having lectures and many other things
10.	and that week was full of interesting things and happy moments	There are many reasons behind why the first week is important at college
11.	I felt happy day by day	and the most important reason is how to go on in a good way to get high marks
12.	I was afraid of how to go on with the new style of studying	many people try to get good marks
13.	but now I am going on in college	and there are people who get accellent degrees,
14.	and it is simple for me.	and they do not learn new things in their college
15.	on one hand the first week is a week of saddness because of the difficult and the different style of having lectures and many other things	So the first week at college is a week of difficulties
16.	but on the other hand it is a week of happiness, first because of getting the college and second because of getting friends and many other things	and that is it
17.	There are many reasons behind why the first week is important at college	
18.	and the most important reason is how to go on in a good way to get high marks	
19.	many people try to get good marks	

20.	and they do hard in their life	
21.	and there are people who get accellent degrees,	
22.	all these point may come from how to be an active one from the first week and how to go on in the following weeks	
23.	sometime the first week at college is a normal week like other weeks special for those who are clever and and smart in their life	
24.	so they see the firs week as a normal week because the are already had more difficult weeks in their life	
25.	while there are students who see the first everyday	
26.	and they do not learn new things in their college	
27.	and those studend see difficulty in their in their life of college, especially those who do not care about knowledge	
28.	so the first week at college is a week of difficulties	
29.	and it is not as simple as normal week.	
30.	and that is it	

The ratio of error-free T-unit /T-unit: 16÷30= 0.5333

2. Example 6.40 (essay number 132): This example can better show the case where no T-unit is correct and only one clause is correct. Nevertheless, when the essay is analyzed using this model of analysis (the various-units-based correctness analysis), the positive achievement of the student is very explicitly shown.

The analysis of this example is detailed but not in the way the previous example is analyzed. I wanted to visualize this in a different way to also show how the coding was actually done on Dexter Coder. The sequences colored in electric blue color are correct (see figure 6.1) and their sum appears in figure 6.2, which is 40 sequences. However, as it is clear from figures 6.3 and 6.4 respectively, no T unit is correct in this essay and only one clause is correct. The

calculation of the accuracy of this example according to the 'various-units based correctness analysis' and its measure (accuracy= CS*MLCS/TWs per essay) is provided in the last paragraph.

Figure 6.1 a screenshot of the essay 132 analyzed using the current method of analysis

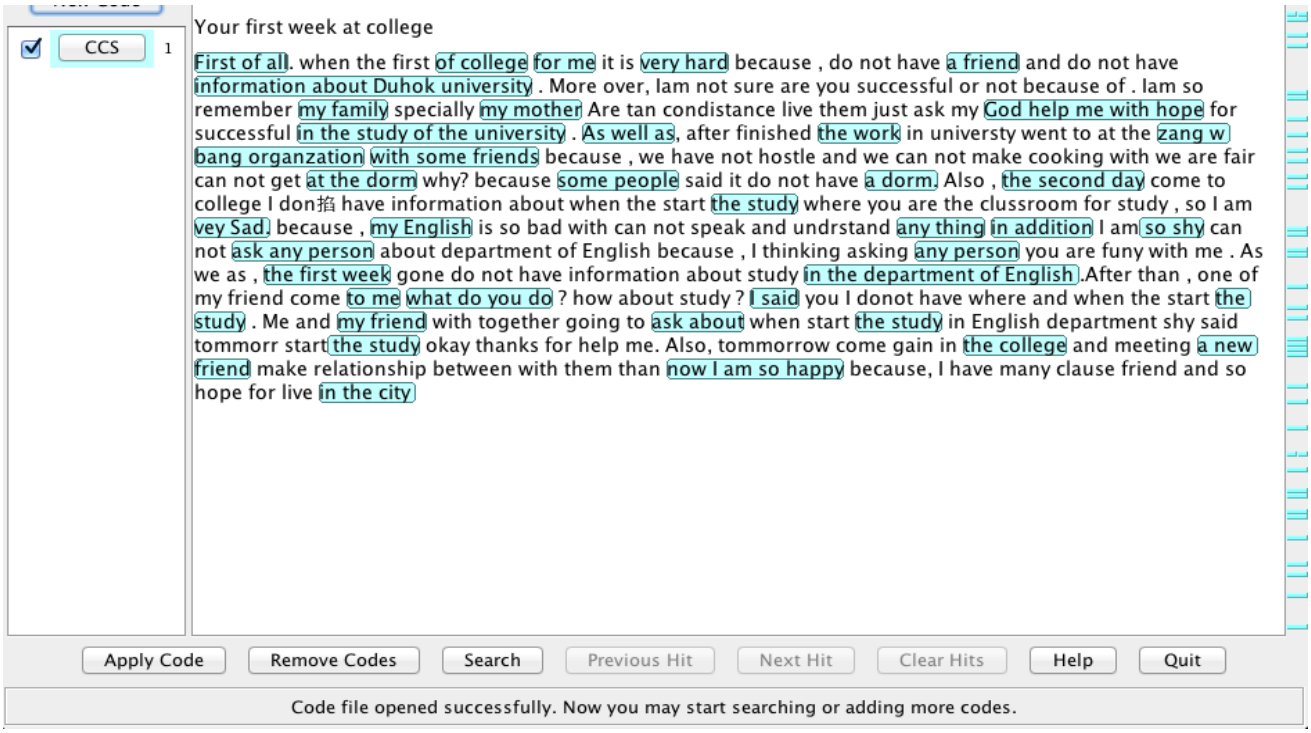
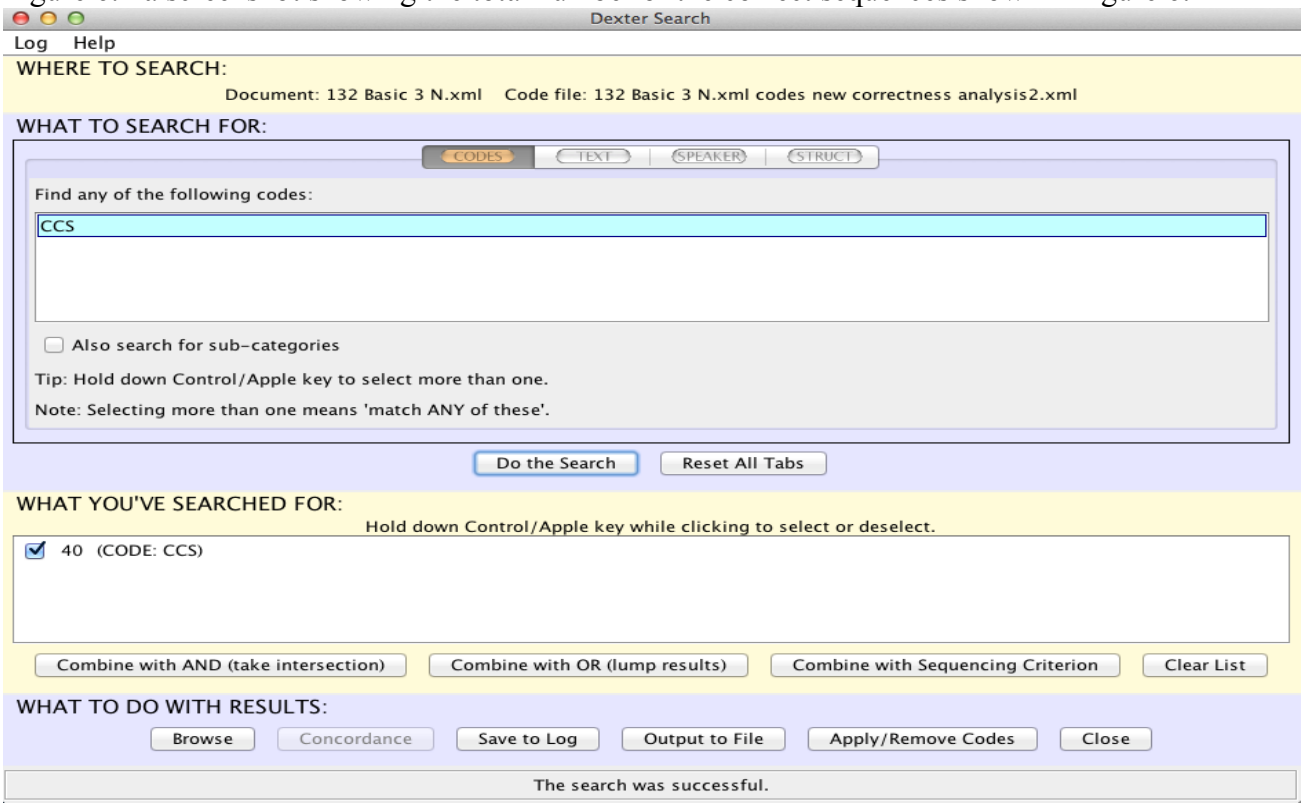
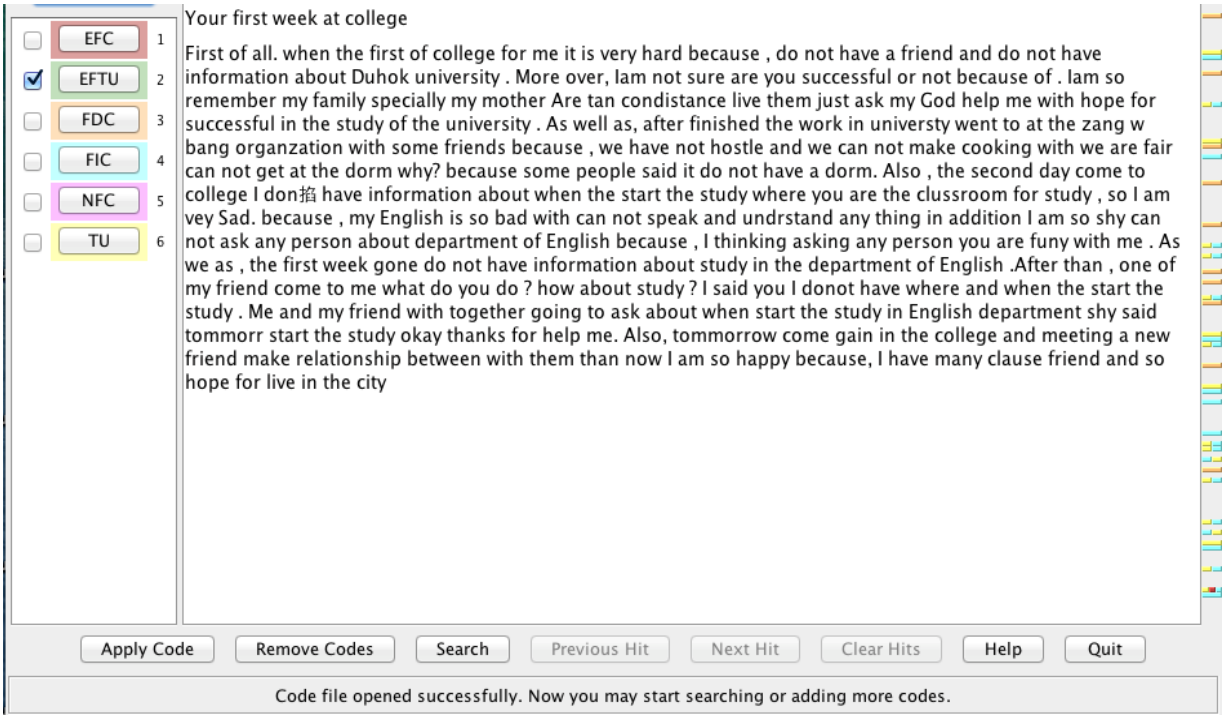


Figure 6.2 a screenshot showing the total number of the correct sequences shown in figure 6.1



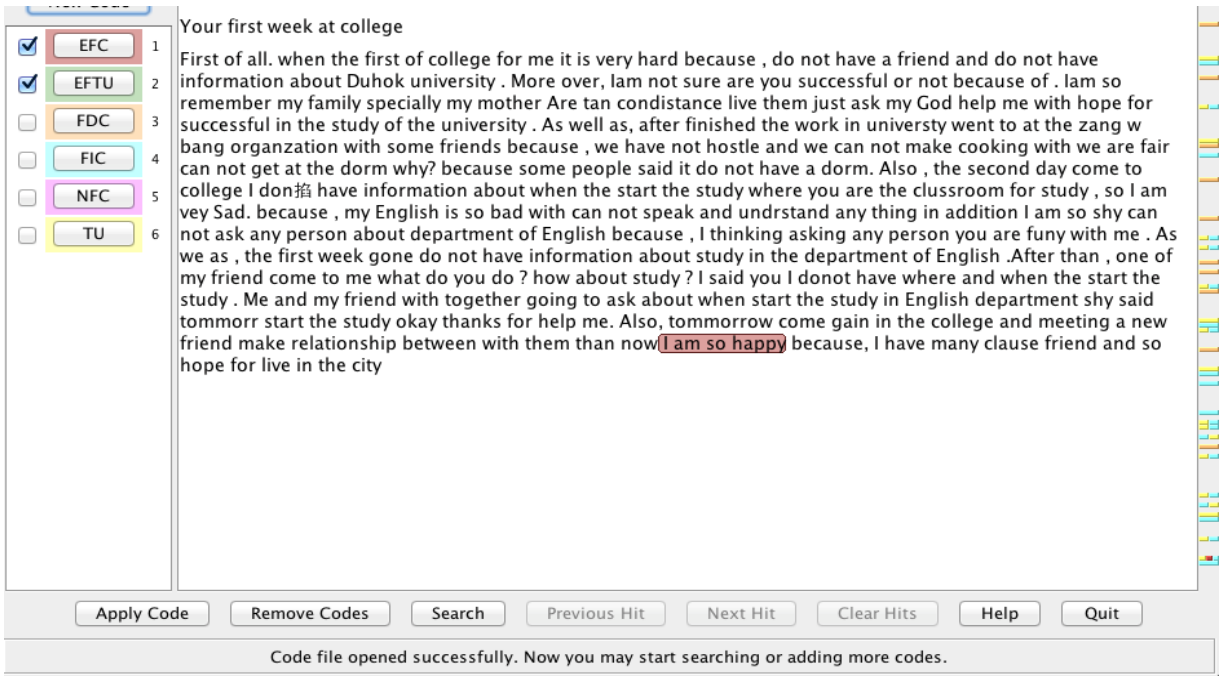
In these two screenshots, it appears that this essay number 132 is divided into 40 correct sequences. The correct sequences are in electric blue color

Figure 6.3 a screenshot showing that none of the T-units in the essay 132 is correct



As obvious from figure 6.3, no T unit is correct in this essay (number 132) as no T unit is colored.

Figure 6.4 a screenshot showing that only one clause in the essay 132 is correct



In this figure 6.4 it appears that only one clause is correct in the whole essay (number 132)

Using the above measure i.e. accuracy= CS*MLCS/TWs per essay, the essay number 132 scored 0.34 (accuracy of essay 132) = 40×(108÷40)÷317= 0.34] but notice that this essay has been ranked as very low in the case of the application of the measures of EFT/T, and EFC/C, (0.000, 0.031 respectively, see appendix J) because it has no correct T-units and one correct clause. From this, one can highlight the difference between the various-units-based correctness analysis and the T-unit and clause-based correctness analysis.

6.4 User engagement

20% of the data that were analyzed with the method of various-units-based correctness analysis have been reanalyzed by another rater, or in other words, a teacher who will be the kind of person who will use the research and who I involved in the rating as a secondary rater. The secondary rating was done in discussion with the first rater (me). The secondary rating yielded sometimes an increase or decrease in the original number of correct sequences. Also, there was an increase or decrease in the number of words (the length) of the sequences. However, the number of correct sequences originally identified in the beginning was not affected to a great extent because even if a part of the sequence was identified by the second rater as wrong, the rest of the sequence would most probably remain as correct. For example, the second rater considered *the beginning* in (e.g. 6.41) *she was like me the beginning* (a sequence produced by the subject number 14) as an incorrect sequence but the sequence *she was like me* remained as correct, and hence the original number of correct sequences was not affected. However, this decreased the length of the sequence by two words (from 6 words to 4 words). A second view was to include *the beginning* as a separate correct sequence, and hence the number of the sequences increased by one with a smaller length. So, the result was two sequences *she was like me* and *the beginning*.

There were some sequences, which I considered correct, but the second rater marked as incorrect. For example, the second rater considered the following sentence as incorrect and provided the correct sequence.

e.g. 6.42 S(10)
the other meals we went to the restaurants (incorrect sequence).
for the other meals we went to the restaurants (correct sequence as provided by the second rater).

Another example of this case is the sentence produced by S(12), e.g. 6.43 *Many people visit Egypt because of pharaoh and their building. The buildings will surprise us if we visit them*

and see how they are big and huge. The second rater considered *Many people visit Egypt because of pharaoh and their building* as incorrect, declaring that it should be *many people visit Egypt because of pharaoh and their building[s]* because the subject of the sentence which immediately follows this sentence starts with the plural head of the noun phrase *the buildings*. An example of the opposite case, which I considered incorrect by mistake and the second rater noticed that they are correct e.g. 6.44 *day by day you will get better* (produced by the subject 16) and the example (6.45) *things are changing* also produced by the subject 16. There were some other cases that raised a lot of discussion between the second rater and myself as the first rater. These cases included whether or not a sequence would be produced by a native speaker. For example, we both discussed whether a native speaker would say we *got into poverty* (e.g.6.46) that was produced by the subject 4.

After a lot of discussion, we both agreed on a new list of numbers of correct sequences for each essay. Then, I correlated the new list of numbers with the numbers that were obtained at the beginning. The result of the correlation was 0.96, although this was not the main aim of the process of user engagement (see section 5.5 in chapter 5 and section 3.6 in chapter 3 for the main aim of user engagement).

6.5 Problems with this method

Although the method worked well, I faced a number of problems while dividing the text into sequences. The following is an explanation of the problems:

The first problem, which could be considered a limitation of the analysis, is that one small error or mistake makes the whole unit incorrect, i.e. error severity is not taken into account. If we had only considered any sequences of words as correct, this problem would have been solved. In the example, e.g. 6.20 S(19) *It was one of my dream,* the sequence of words *it was one of my* is correct. If only the sequential view of language was taken into account, it would have been straightforward and easy to say this sequence is correct. However, as we have considered structured units, it is not possible here to take *it was only one of my* as a one structured unit because it is only a sequence of words. Hence, the whole unit *one of my dream* is considered to be wrong because of the absence of the pluralization marker that should have been added to the word *dream.*

The second problem is that despite the fact that a number of explicit criteria have been provided for identifying the correct units (sequences) and their boundaries, it is difficult to set

a clear definition for the unit because all units are of different lengths and structures. This made it difficult to be quite consistent while breaking the text down into correct units. The example (6.22) above *She said to us that the system of education* very well exemplifies this case as it is not easy to decide whether to separate *she said to us that* and *the system of education* or take the whole sequence together.

e.g. 6.47 S(7)
an agreement to make a party at college to know other students.

Example 6.47 is another example that raises the question of whether to take the whole sequence together or separately. Nevertheless, this problem does not affect the measure because it involves both the total number and the mean length of correct sequences.

The word order is the third problem. Consider the following example (6.48)
e.g. 6.48 S(148) *my husband bought for me a beautiful ring.*

Now taking this sentence together it has a word order problem but if the constituents are taken separately, they are accurate and well-formed.

The fourth problem is that although I have tried very hard with the criteria to exclude subjectivity, subjectivity does come in to a certain extent. Consider the following examples, which are borderline between being correct and being incorrect.

e.g. 6.49 S(143)
it will be a forgettable experience for our childhood. That I am sure they will never forget because they have seen how their parents were taking

In this example, the underlined part is grammatically correct but lexically it is not. This is because *experience* and *forgettable* are not likely to collocate as an experience is an experience because you cannot forget it. It was not easy to decide about such cases whether they are erroneous or error-free.

Finally, it was not easy sometimes to decide whether a native English speaker will possibly produce a given sentence, as it seems to be a direct translation of the L1. I noticed this because I share the two L1s spoken by the subjects (Arabic and Kurdish).

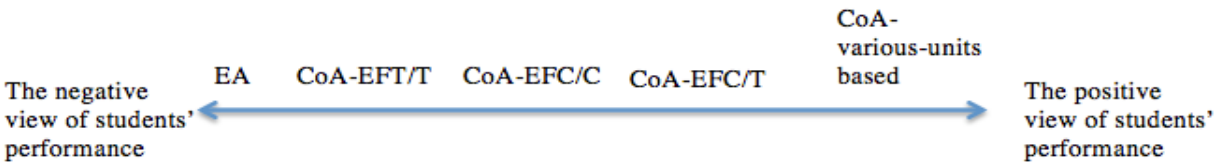
e.g. 6.50 S(61) *I always metion their names to God.*
e.g. 6.51 S(62) *they removed a part of the veil on my eyes.*

These two sentences seem to be a direct translation from Kurdish. The sentence in example 6.50 means *I always pray for them* and the example 6.51 means *Now I could see the truth*.

6.6 Results and discussion

The same 80 essays that have been analyzed in the previous chapter have been re-analyzed using this method of various-units-based correctness analysis. This step was taken in order to show how looking at students' performance from different perspectives leads to different views of their language. T-unit analysis involves taking relatively long sequences and assessing them as correct only if they have no errors at all. As a result, there is a low likelihood of correctness. This might be termed the 'negative view' of language proficiency. In various-units-based correctness analysis, on the other hand, correctness is assumed to attach to all sequences, including sub-clausal ones. Correct sequences may therefore be short or long. This might be termed the 'positive view'. Even the very 'low level' essays in my corpus that have no correct T-units or clauses have scored higher when they are subjected to analysis using this method. I have attempted to raise the profile of the positive view of language learning by moving from looking only at errors in chapter five (the full negative view) to looking at the smallest *correct* stretches in this chapter (the full positive view). If one puts these two views (the erroneousness view and the correctness view) at two ends of a spectrum, error analysis will be at the erroneousness end but the method of various-units-based correctness analysis will be at the correctness end. Figure 6.5 shows this spectrum

Figure 6.5 a spectrum of the analytical views of language



The results of applying the method of various-units-based correctness analysis to the 80 essays are tabulated in tables 6.3, 6.4, 6.5 and 6.6. Each table consists of 6 columns. The first column contains the students' codes. The second column indicates the number of correct sequences; the third column indicates the average length of these correct sequences. The fourth column contains the results of multiplying columns two and three. The fifth column shows the total words in the essay and the last column shows the result of dividing the numbers in columns four and five. The numbers in this column (column 6) are between 0 and 1, and they are measures of accuracy in students' essays.

Table 6.3 the results of applying the method of various-units-based correctness analysis to 20 essays written by 20 third year students/ school of Arts.

students' codes	Number of correct sequences (CS)	mean length of correct sequences (MLCS)	(CS)*(MLCS)	total words in essay	(CS)*(MLCS)/ total number of words in essay
1A3N	50	6.56	328	380	0.86
2A3N	63	5.16	325	454	0.72
3A3N	63	6.32	398	524	0.76
4A3N	52	7.54	392	477	0.82
5A3N	84	3.62	304	584	0.52
6A3N	72	6.85	493	577	0.85
7A3N	46	5.15	237	354	0.67
8A3N	44	6.23	274	380	0.72
9A3N	39	9.31	363	453	0.80
10A3N	61	7.62	465	554	0.84
11A3N	74	3.38	250	605	0.41
12A3N	53	8.47	449	504	0.89
13A3N	49	9.84	482	516	0.93
14A3N	54	6.80	367	513	0.72
15A3N	44	4.02	177	312	0.57
16A3N	53	5.57	295	457	0.65
17A3N	51	5.22	266	347	0.77
18A3N	60	4.85	291	500	0.58
19A3N	74	5.72	423	583	0.73
20A3N	33	15.39	508	538	0.94

Table 6.4 the results of applying the method of various-units-based correctness analysis to 20 essays written by 20 fourth year students/ school of Arts.

Students' codes	Number of correct sequences (CS)	Mean length of correct sequences (MLCS)	(CS)*(MLCS)	Total words in essay	(CS)*(MLCS)/ total number of words in essay
58A4N	24	9.42	226	275	0.82
59A4N	48	8.29	398	462	0.86
60A4N	46	8.76	403	492	0.82
61A4N	65	5.00	325	419	0.78
62A4N	59	8.66	511	608	0.84
63A4N	57	6.82	389	475	0.82
64A4N	73	5.14	375	535	0.70
65A4N	93	4.80	446	622	0.72
66A4N	57	6.37	363	521	0.70
67A4N	81	5.02	407	708	0.57
68A4N	72	4.13	297	461	0.64
69A4N	60	4.03	242	465	0.52
70A4N	38	9.95	378	408	0.93
71A4N	66	5.08	335	561	0.60
72A4N	50	5.48	274	446	0.61
73A4N	70	6.97	488	633	0.77
74A4N	88	4.01	353	752	0.47
75A4N	49	4.43	217	383	0.57
76A4N	68	6.68	454	531	0.85
77A4N	40	7.88	315	380	0.83

Table 6.5 the results of applying the method of various-units-based correctness analysis to 20 essays written by 20 third year students/ school of Basic Education

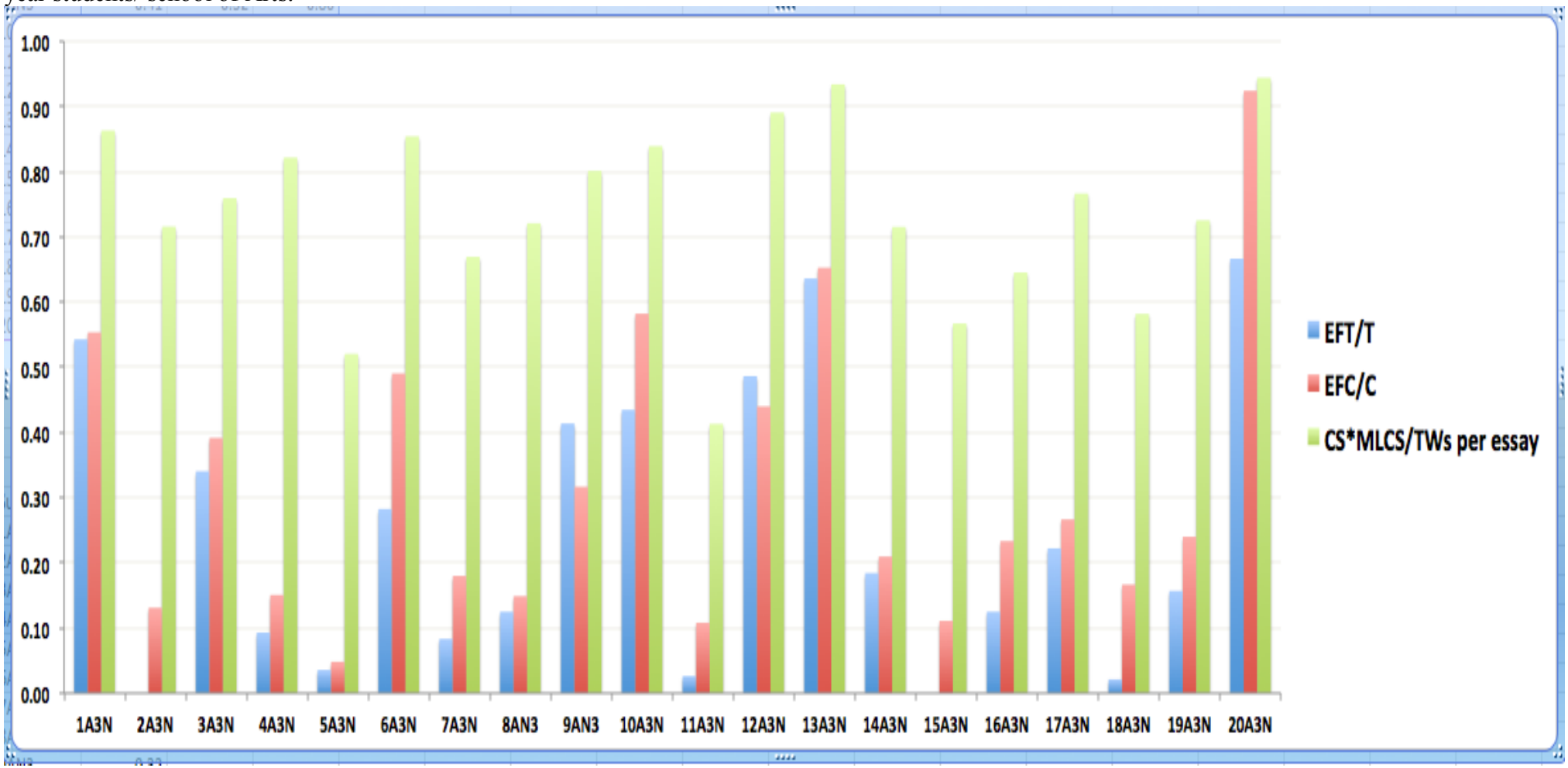
Students' codes	Number of correct sequences (CS)	Mean length of correct sequences (MLCS)	(CS)*(MLCS)	Total words in essay	(CS)*(MLCS)/ total number of words in essay
113B3N	41	5.83	239	344	0.69
114B3N	22	10.68	235	258	0.91
115B3N	40	7.95	318	401	0.79
116B3N	42	8.43	354	390	0.91
117B3N	65	5.31	345	439	0.79
118B3N	41	5.07	208	335	0.62
119B3N	36	4.67	168	258	0.65
120B3N	37	4.95	183	301	0.61
121B3N	75	5.28	396	578	0.69
122B3N	66	8.12	536	618	0.87
123B3N	70	3.90	273	489	0.56
124B3N	27	7.33	198	214	0.93
125B3N	80	4.46	357	521	0.69
126B3N	67	3.40	228	397	0.57
127B3N	32	5.50	176	221	0.80
128B3N	41	3.07	126	275	0.46
129B3N	21	3.90	82	167	0.49
130B3N	76	3.96	301	530	0.57
131B3N	44	4.32	190	297	0.64
132B3N	40	2.70	108	317	0.34

Table 6.6 the results of applying the method of various-units-based correctness analysis to 20 essays written by 20 fourth year students/ school of Basic Education.

Students' codes	Number of correct sequences (CS)	Mean length of correct sequences (MLCS)	(CS)*(MLCS)	Total words in essay	(CS)*(MLCS)/ total number of words in essay
139B4N	70	5.94	416	528	0.79
140B4N	57	7.42	423	482	0.88
141B4N	51	5.76	294	451	0.65
142B4N	39	3.92	153	272	0.56
143B4N	62	4.39	272	471	0.58
144B4N	51	3.22	164	392	0.42
145B4N	24	6.67	160	201	0.80
146B4N	37	5.95	220	257	0.86
147B4N	27	11.19	302	343	0.88
148B4N	64	5.33	341	483	0.71
149B4N	81	4.58	371	537	0.69
150B4N	53	5.64	299	404	0.74
151B4N	33	6.48	214	279	0.77
152B4N	58	5.19	301	449	0.67
153B4N	39	5.74	224	346	0.65
154B4N	71	4.20	298	581	0.51
155B4N	27	10.81	292	308	0.95
156B4N	81	4.63	375	564	0.66
157B4N	50	4.52	226	351	0.64
158B4N	61	8.13	496	601	0.83

Figure 6.6 is a comparison of (1) the results of applying the method of correctness analysis used in the previous chapter (the T-unit- and clause-based correctness analysis) and its two measures, (EFT/T) and (EFC/C) to the essays written by the third year students/ school of Arts, with (2) the results of applying the method of various-units-based correctness analysis that is used in this chapter and its measure (i.e. the total number of correct sequences × the mean length of the correct sequences ÷ total number of words per essay (CS*MLCS/TWs per essay) to the same sample of essays. The horizontal axis shows the codes of the students. For example 1A3N= subject number 1 from school of Arts (A), third year student (3) who wrote a narrative essay (N). The vertical axis shows the results of applying the three measures to these students' essays.

Figure 6.6 the results of applying the method of various-units-based correctness analysis with its measure (CS*MLCS/TWs per essay) compared to the results of applying the method of T-unit and clause-based correctness analysis with its two measures EFT/T and EFC/C to 20 essays written by the third year students/ school of Arts.



The three measures measure correctness on a scale of 0–1. In other words, if a student produces no correct T-units in the case of EFT/T, correct clauses in the case of EFC/C and correct sequences in the case of CS*MLCS/TWs per essay, they will score 0 and if a student produces all the T-units, clauses or sequences correctly, they will score 1. One difference between the measure of various-units-based correctness analysis (CS*MLCS/TWs per essay) and the other two measures (EFT/T and EFC/C) is that in the various-units-based correctness analysis the mean length of correct sequences (units) is involved in the calculation of students' achievement. This is because the units are of different lengths that start from a combination of two words to a high number, especially in the case of compound complex sentences. Accordingly, a student who has produced a small number of correct sequences but with a high length will score closer to 1 than a student who has produced a higher number of very short correct sequences. Notice the difference between subjects: 5A3N and 20A3N in table 6.3, 65A4N and 70A4N in table 6.4, 125B3N and 114B3N in table 6.5, 156B4N and 145B4N in table 6.6. Although subjects 5A3N, 65A4N, 125B3N, and 156B4N have produced more correct sequences (84, 93, 80, 81 respectively) than subjects 20A3N, 70A4N, 114B3N and 145B4N respectively (33, 38, 22, 24 respectively), their overall accuracy is less (0.52 for subject 5A3N, 0.72 for subject 65A4N, 0.69 for subject 125B3N, 0.66 for 156B4 compared to 0.94 for 20A3N, 0.93 for 70A4N, 0.91 for 114B3N, 0.80 for 145B4N respectively). What changed the results is the mean length of the units (3.62 for subject 5A3N, 4.80 for subject 65A4N, 4.46 for subject 125B3N, 4.63 for 156B4 compared to 15.39 for 20A3N, 9.95 for 70A4N, 10.68 for 114B3N, 6.67 for 145B4N respectively).

Moreover, it is very clear from figure 6.6 that the measure CS*MLCS/TWs gives more credit to the students. A student who has produced no correct T-units and only very few correct clauses has scored high when their performance is evaluated using the measure (CS*MLCS/TWs per essay). Very good examples of this case are the students 2A3N, 4A3N, 5A3N, 7A3N, 11A3N, 15A3N, and 18A3N in figure 6.6. These students scored below 0.10 in the case of the measure EFT/T (2A3N and 15A3N have not produced even one error-free T-unit) but have scored above 0.50 when their performance was evaluated using the measure (CS*MLCS/TWs per essay). As for the measure EFC/C, despite the fact that these students have scored a little higher (above 0.20) than in the case of EFT/T, they still show a very significant difference to the measure (CS*MLCS/TWs per essay). If one takes the student 2A3N as an example, one can very clearly notice this difference. This student's essay scored very low on the two measures EFT/T, EFC/C, (0.000, 0.13, respectively) but much higher in the case of applying various-units-based correctness analysis (0.72). To put it another way,

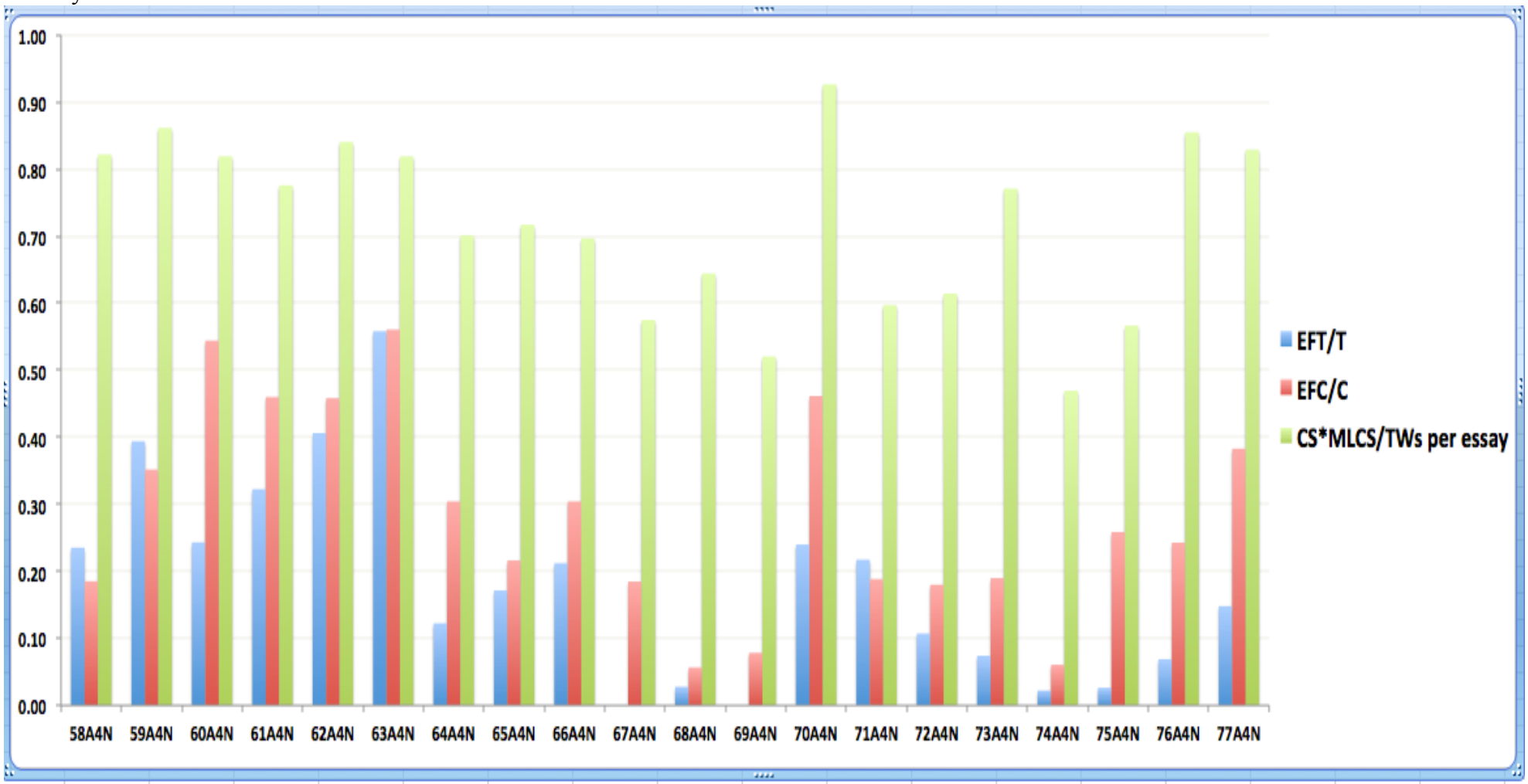
this student has produced no correct T-units and only 8 error-free clauses (see appendix J) but 63 correct sequences with a mean length 5.16 (see appendix K and table 6.3 above). Even with the measure EFC/T that evaluates students' performance on a scale above 1, this student has scored much less (0.33) than when his/her performance is assessed with this measure. Based on this discussion, it is worth mentioning that the aim of finding this measure is to have a measure that benefits the students but which is also fair to the students or a measure, which can discriminate finely between the levels of students.

This discussion indicates that this method accounts for correctness rather than erroneousness. In other words, it measures accuracy more in terms of 'achievement' than in terms of 'failure'. It also shows, firstly, that the smaller the unit, the less prone it is to errors. Secondly, considering one strictly defined unit like a clause or a T-unit may limit the amount of credit given to students for writing correct English. On the actual performance level, this signifies two important points. The first one is that the students have the ability to produce more correct short units than correct long units. In other words, students face difficulty in putting the small units, whether words, phrases or multi-word units together to make longer correct structured stretches of language. This might be explained by the fact that language is probably stored in the form of small 'bits and pieces' in the students' brains without them having enough knowledge about the syntax of relating these 'bits and pieces' to each other to produce longer units. The second is that one simple error excludes a whole long unit (T-unit or a clause) from the correct performance, leading to most of the performance becoming incorrect.

What is quite visible in this figure 6.6 as well is the consistency of students' scores with all the three measures. Although the actual values are much higher with some measures than others, the pattern of scoring remains virtually consistent for most of the students. Examples of exceptions to this consistency are the students 12A3N and 1A3N, where 1A3N scored higher than 12A3N on the two measures EFT/T, EFC/C but 12A3N scored slightly higher on the measure (CS*MLCS/TWs per essay); the same case applies to students 3A3N and 4A3N. On the other hand, a good instance of the consistency is that the student number 20A3N was the best among all students with the two measures EFT/T, EFC/C and he/she remained so with the measure (CS*MLCS/TWs per essay). The next best achievement was attained by the student 13A3N on the two measures EFT/T, EFC/C, keeping the same ranking position as the second best student with the measure (CS*MLCS/TWs per essay). The performance of the students 16A3N, 17A3N, 18A3N, and 19A3N also exemplifies well the consistency in the pattern of scoring along the three measures.

Figure 6.7 is a comparison of (1) the results of applying the method of correctness analysis used in the previous chapter (the T-unit- and clause-based correctness analysis) and its two measures, (EFT/T) and (EFC/C) to the essays written by the fourth year students/ school of Arts, with (2) the results of applying the method of various-units-based correctness analysis that is used in this chapter and its measure (i.e. the total number of correct sequences × the mean length of the correct sequences ÷ total number of words per essay (CS*MLCS/TWs per essay) to the same sample of essays. (For more information about the abbreviations, see figure 6.6).

Figure 6.7 the results of applying the method of various-units-based correctness analysis and its measure (CS*MLCS/TWs per essay) compared to the results of applying the method of T-unit- and clause-based correctness analysis with its two measures EFT/T and EFC/C to 20 essays written by the fourth year students/ school of Arts



As in figure 6.6, in figure 6.7 the students 67A4N, 68A4N, 69A4N, 73A4N, 74A4N, 75A4N, and 76A4N present a good example of the case of very low achievement on the measure EFT/T (below 0.10, the students 67A4N, 69A4N even scored 0 on this measure) and much higher achievement (above 0.50) on the measure (CS*MLCS/TWs per essay). As the case with student 2A3N in figure 6.6, student 76A4N is a very good example of a very low achievement on the measure EFT/T (0.04) and a very high achievement (0.85) on the measure (CS*MLCS/TWs per essay). This student has produced only 1 T-unit correctly out of a total of 28 T-units (see appendix J) but 68 correct sequences with a mean length 6.68 when evaluated by the method (CS*MLCS/TWs per essay) (see table 6.4 and appendix K). Concerning the students' performance on the measure EFC/C, although it is higher than in the case of the measure EFT/T, it still remained much lower than their performance on the measure (CS*MLCS/TWs per essay).

The consistency of the scoring on the three measures is not as remarkable as the case with the three measures in figure 6.6. Student 63A4N is ranked first on the two measures EFT/T and EFC/C but on the measure (CS*MLCS/TWs per essay) the student 70A4N scored higher. This variation in the scoring on the three measures is not only restricted to these two students but one can see the variation in most students' performance (notice the difference between students 59A4N and 60A4N, 65A4N and 66A4N, 76A4N and 77A4N). Nevertheless, it is true that there is variation from one student to another as far as their scores on the three measures are concerned but the variation seems to be low.

This little variation might be ascribed to the mean length of the correct sequences that have been identified using the method of various-units-based correctness analysis. Considering the two students 63A4N and 70A4N, this becomes clear. What made the student 70A4N score higher on this method is the longer *correct* sequences he/she has produced (mean length of *correct* sequences = 9.95 while the mean length of the sequences produced by 63A4N = 6.82). The reason why the student 70A4N scored less on the measure EFT/T is the longer T-units produced by him/her (mean length of T-unit = 16.36 for this student while the mean length of the T-units produced by the student 63A4N = 11.04). Also, most of the T-units of this student have been marked as incorrect because of the punctuation confusion error. These two reasons made the performance of the student 70A4N on T-units more prone to errors. The point I am trying to make here is that possibly one long T-unit has been rendered incorrect because of one small error or mistake but with the various-units-based correctness analysis that T-unit has been divided into two long or three correct sequences. The following example from student 70A4N's essay very well exemplifies this point.

e.g. 6.52 S(70)

Although meeting various friends is not a matter of exploitation or having fun all the time, but it is a process of cooperation and helping each other in all situations (one T-unit)

In this example, the use of *but* has made the whole T-unit incorrect. However, when analyzing this essay using the various-units-based correctness analysis, this T-unit has been divided into two long correct sequences: *Although meeting various friends is not a matter of exploitation or having fun all the time* (one correct sequence) and *it is a process of cooperation and helping each other in all situations* (one correct sequence).

Another example of this case (from another student) is the following:

e.g. 6.53 S(58)

Sometimes life strikes us with its utmost power and compels us to dive in harsh tribulations and fill up our spirit with experiences. (one T-unit)

In this example, the whole T-unit is rendered incorrect just because of the absence of the third person singular marker (-s); but, when analyzing this sentence with the current method, most of the language in the unit is taken as correct:

S(58) *Sometimes life strikes us with its utmost power and compels us to dive in harsh tribulations and fill up our spirit with experiences.*

Another example that may confirm this is the performance of the student 63A4N compared to the performance of the student 62A4N on the three measures. We can observe that student 63A4N has scored higher than the student 62A4N on the T-unit and clause measures but lower on the various-units-based correctness analysis method (CS*MLCS/TWs per essay). This is again because of the difference in the mean length of the correct sequences identified (6.82 for 63A4N compared to 8.66 for 62A4N). But the mean length of T-unit in the case of student 62A4N is higher (mean length of T-unit =19) than the mean length of the T-units produced by the student 63A4N (11.04), which makes the production of the student 62A4N in terms of T-units more likely to contain errors.

Table 6.7 the results of the measure (CS*MLCS/TWs per essay) applied to the essays of the third and fourth year students/ school of Arts (20 essays per each year)

	0.00- 0.10	0.11-0.20	0.21-0.30	0.31-0.40	0.41-0.50	0.51-0.60	0.61-0.70	0.71-0.80	0.81-0.90	0.91-1
3rd year Arts	0	0	0	0	1	3	2	7	5	2
4th year Arts	0	0	0	0	1	4	4	3	7	1

Comparing the performance of the third year students to that of the fourth year students, it is obvious from table 6.7 that they are exactly the same until the range 0.41–0.50 as none of the third and fourth year students have scored within the range 0.00–0.40 and only 1 student in each group scored 0.41–0.50. The difference only starts at the range 0.51–0.60. Again, as is the case with the other measures (EFT/T, EFC/C, EFT/C), although the third year group has two students who have scored close to 1 (above 0.90)[2], the fourth year group seem to have more students who scored between 0.81–0.90 (7 students compared to 5 from the third year group). However, more students from the third year group scored 0.71-0.80 (7 third year students compared to 3 fourth year students), and more fourth year students scored 0.61 - 0.70 (4 compared to 2 in the third year group) and between 0.51-0.60 (4 compared to 3 third year students).

Two important points are worth mentioning here. One is that no student in either group scored below 0.40 and only one student in each group scored between 0.41–0.50. This explains why I have described this method as representing the positive view of language learning (figure 6.5). The second point is that the two groups mostly differ in the second digit of the decimal number. For example, for those who scored between 0.81 and 0.90 in the fourth year group, their scores vary between 0.82 and 0.86 but the scores of the third year group vary between 0.82 and 0.89. However, the student who scored between 0.41–0.50 in the fourth year scored much higher (0.47 compared to 0.41) than that in the third year, who scored within this range. Thus, there does not seem to be a great difference between the performance of the two groups, though the third year students appear to have scored better because two of them scored close to 1 (above 0.90) and their 0.80s range between 0.82–0.89 (compared to the 0.80s of the fourth year students, which range between 0.82–0.86). In addition to that, more third year students scored between 0.71–0.80 than the fourth year students. Although, there is a small difference, this difference is not significant statistically (P value = 0.66). Comparing the two groups in terms of the measures EFT/T, EFC/C, the fourth year students looked to have done better overall (see chapter 5 for a discussion of this point). This difference in the level of students from one measure to another might be due to the variation in the levels of the fourth year students across different measures.

Figure 6.8 is a comparison of (1) the results of applying the T-unit- and clause-based correctness analysis used in the previous chapter and its two measures, (EFT/T) and (EFC/C) to the essays written by the third year students/ school of Basic Education, with (2) the results of applying the method of various-units-based correctness analysis that is used in this chapter and its measure (i.e. the total number of correct sequences × the mean length of the correct sequences ÷ total number of words per essay (CS*MLCS/TWs per essay) to the same sample of students. The horizontal axis shows the codes of the students. For example 113B3N= subject number 113 from school of Basic Education (B), third year student (3) who wrote a narrative essay (N). The vertical axis shows the results of applying the three measures to these students' essays.

Figure 6.8 the results of applying the method of various-units-based correctness analysis with its measure (CS*MLCS/TWs per essay) compared to the results of applying the methods of T-unit and clause-based correctness analysis with its two measures EFT/T and EFC/C to 20 essays written by the third year students/ school of Basic Education

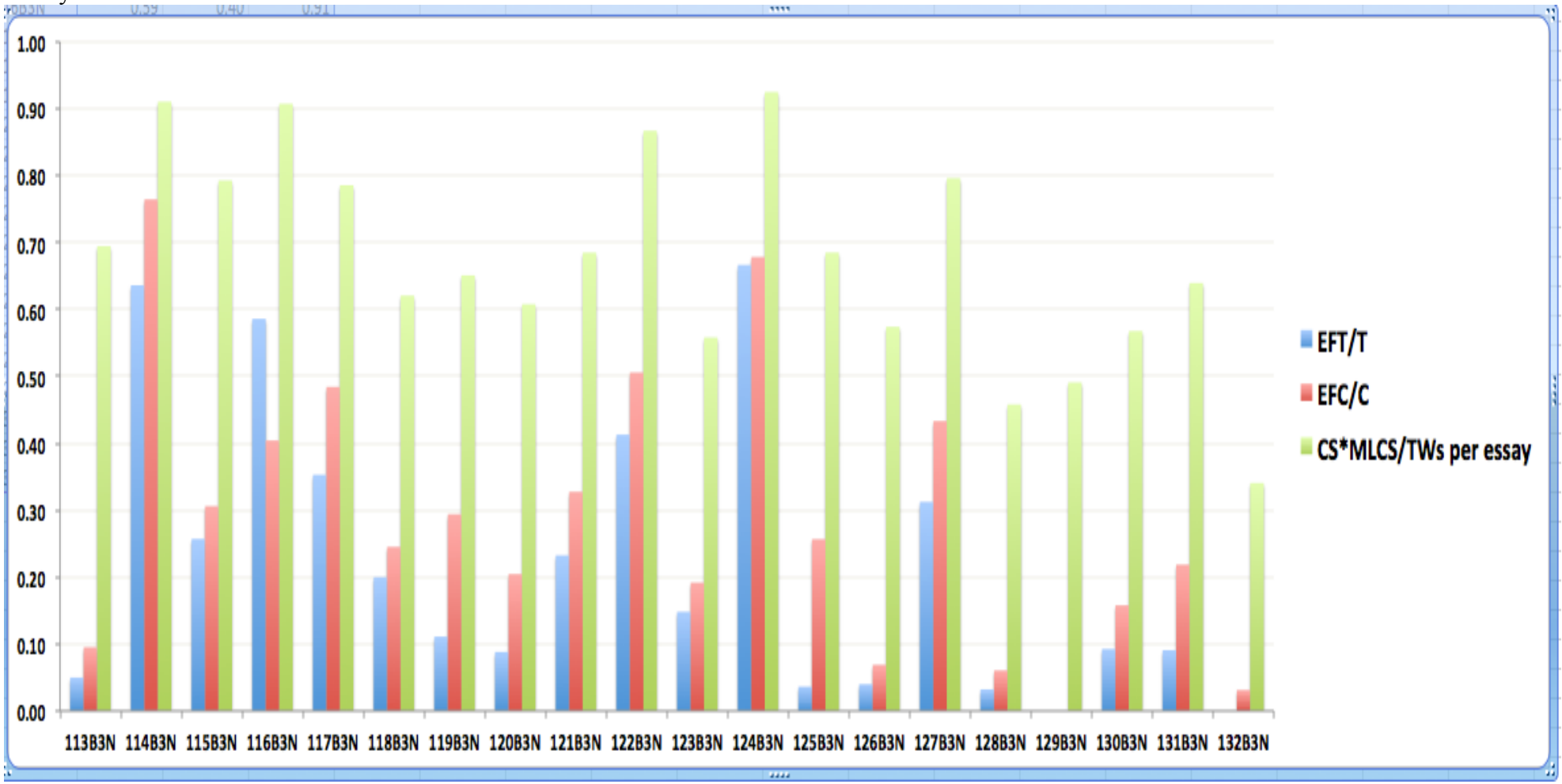
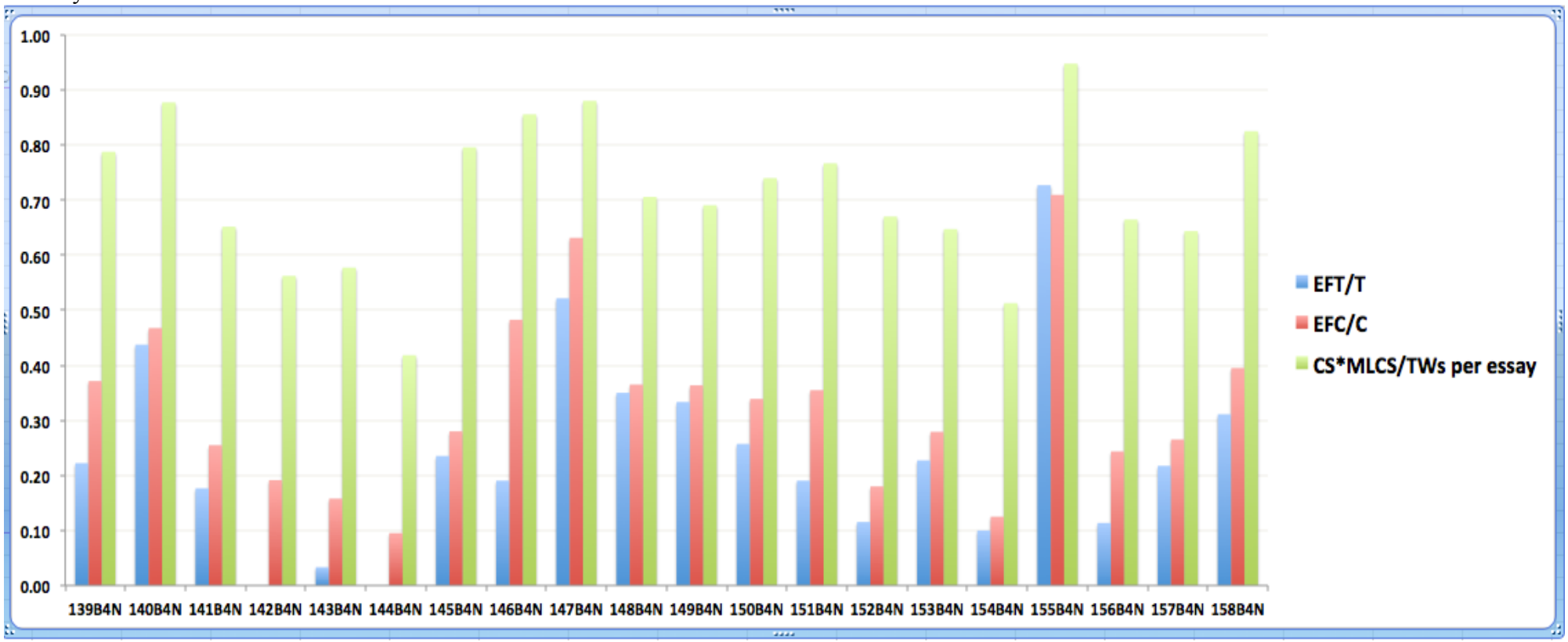


Figure 6.9 is a comparison of (1) the results of applying the method of T-unit- and clause-based correctness analysis used in the previous chapter and its two measures, (EFT/T) and (EFC/C) to the essays written by the fourth year students/ school of Basic Education, with (2) the results of applying the method of various-units-based correctness analysis that is used in this chapter and its measure (i.e. the total number of correct sequences × the mean length of the correct sequences ÷ total number of words per essay (CS*MLCS/TWs per essay) to the same sample of students. For more information about the abbreviations, see the description of the figure 6.8).

Figure 6.9 the results of applying the method of various-units-based correctness analysis with its measure (CS*MLCS/TWs per essay) compared to the results of applying the method of T-unit- and clause-based correctness analysis with its two measures EFT/T and EFC/C to 20 essays written by the fourth year students/ school of Basic Education



As with the previous figures (6.6 and 6.7), these two figures as well (figures 6.8 and 6.9) show that the students obtained higher scores when their language was analyzed with the measure (CS*MLCS/TWs per essay) than with the other two measures (EFT/T, EFC/C), and they are consistent in their performance on the three measures i.e. the students who have scored high on the measures EFT/T and EFC/C have also scored high on the measure (CS*MLCS/TWs per essay) except for few cases such as the student 151B4N in figure 6.9. This student has scored lower than the students 149B4N and 150B4N in the same figure on the measures EFT/T and EFC/C but higher on the measure (CS*MLCS/TWs per essay).

Table 6.8 the results of the measure (CS*MLCS/TWs per essay) applied to the essays of the third and fourth year students/ school of Basic Education (20 essays per each year)

	0.00- 0.10	0.11-0.20	0.21-0.30	0.31-0.40	0.41-0.50	0.51-0.60	0.61-0.70	0.71-0.80	0.81-0.90	0.91-1
3rd year Basic Education	0	0	0	1	2	3	7	3	1	3
4th year Basic Education	0	0	0	0	1	3	6	5	4	1

Comparing the achievement of the third year group/school of Education to that of the fourth year group of the same school (table 6.8), one can see that the students in both groups scored the same on the score between 0.00–0.30 as none of the students falls within this range. The difference starts at the score 0.31–0.40 as only one student of the third year grade is within this category. Although three third year students outperformed the fourth year students by scoring above 0.90, considerably more fourth year students scored between 0.81–0.90 (4 fourth year students compared to 1 third year students). Also, on the level of 0.70s the fourth year students outperformed the third year students (5 fourth year students compared to 3 third year students). It is also noteworthy that none of the fourth year students scored between 0.31-0.40 but one third year student did, meaning that even with the 'achievement' measure, this student could not reach the level of 0.41. To sum up the results of the comparison, I can say that both groups differ from each other in their performance on different levels. However, this difference is not statistically significant (P=0.50). It is, though, fair to say that overall the fourth year students slightly outperformed the third year students on more levels.

6.7 Conclusion

This chapter has examined second language writing from a new angle. It represents an important attempt to measure accuracy in learner language according to a 'positive point of view' rather than a 'negative point of view'. Researchers have already dealt with this issue but in a different way. They have carried out the 'single-type-unit' analysis, taking only one type of unit at one time. For example, they have measured accuracy using T-units, clauses, and sentences, AS-units etc. No research to date has looked at correct sequences of any unit type. When one specifies the unit, especially if it is a long unit, the learner's production is not well measured. This is because the researcher is obliged to take only one type of unit and it has to be only that type of unit and it has to be completely correct. This lessens the possibility of finding many of these units because even one small error (which might be even a 'slip of the pen' mistake) will make a whole unit incorrect, even though it may contain substantial error-

free language. The learner in this case loses credit for their correct language. To maximize the identification of correct language performance, I have used a different method of correctness analysis calling it the 'various-units-based correctness analysis' because it allows for considering different types of units even sub-clausal units.

The results show that all students, with no exception, scored higher on this measure that on the T-unit and clause -based measure because it is more flexible and considers most of the correct language of the students. It was, however, not easy to set a list of criteria for identifying units. What caused a particular difficulty was the use of many types of units. It is certainly easier to divide a text into one kind of unit than many because the researcher can more easily set criteria and definitions. That is why I have stated these problems in one section (section 6.5).

The results also show that the students kept their levels across the three measures i.e. EFT/T, EFC/C and (CS*MLCS/TWs per essay). In other words, the students who scored high on the EFT/T, EFC/C measures scored high on this measure as well. One exception to this point was the performance of fourth year students/ school of Arts, which is attributed in most cases to the length of T-units and clauses and the mean length of the correct units considered for the measure (CS*MLCS/TWs per essay).

As for comparing the performance of levels to each other, it can simply be said that the students differed but not to a great extent when the achievement of the third year students/ school of Arts was compared to that of the fourth year students/ the same school. Nevertheless, the third year students seemed to have scored higher than the fourth year students, which is unlike the results in chapter five. This can be due to the fact that the fourth year students were inconsistent in their achievement on the three measures (EFT/T, EFC/C and CS*MLCS/TWs per essay). Concerning the performance of the third and fourth year students of the school of Basic Education, it is clear that the fourth year outperformed the third year, meaning the results of the comparison remained the same as in chapter five.

I would argue that this method has worked better than evaluating students' performance in terms of T-units and clauses. Most of the scores attained corresponded to my intuitive holistic evaluation as an assessor, researcher and teacher of writing. This is because students' performance has been evaluated on 'various correct units' basis, meaning no account has been taken of what type the units are. In the case of T-units and clauses, the possibility of correctness decreases because the correct unit has to be a T-unit or a clause. In other words,

this method works more than the T-unit- and clause-based measures on the principle of 'achievement rather than failure' that is mentioned above.

Finally, although this method worked well, the students are given more credit, and most of their correct performance is accounted for, one of the problems with it is that it does not take the severity of error into consideration and another problem, as mentioned earlier, is that it is not easy to divide the text into different units. Having considered (1) these two points, (2) the weak points of the T-units- and clause-based correctness analysis, and (3) the importance of considering error analysis besides correctness analysis, combining these methods (i.e. the T-units- and clause-based correctness analysis, the 'various-units-based' correctness analysis, and error analysis) in one method is an appropriate course of action. This is (1) not to leave any pieces of language unanalyzed, whether correct pieces in the case of error analysis or erroneous pieces in the case of correctness analysis, (2) to ensure that the low level students who wrote very bad quality writing are given credit for having produced a number (even though that may be small) of correct sequences of language regardless of how big or small these sequences are, (3) to consider error type and gravity so that a fair distinction is made between the erroneous units and students are privileged to have the severity of their errors judged, and (4) not to allow too much of the language to be correct. I have called this method 'An Integrated Approach to Achievement' (IAA), and it is the focus of the next chapter.

Notes

1. Out of context this sentence may not make sense but in this context it does. The system of teaching in the Iraqi Kurdistan universities is either 'annual' or courses (semesters). 'Annual' means that students' work is assessed throughout the whole year and the student gets his/her final mark at the end of the year but 'courses' means the students' work is assessed two times during the academic year and when the students finish one course (semester), they start another course with different modules and their assessment is independent of semester one.
2. They are the same students who have been to an English speaking country or learned their English in a different context (student 20A3N and student 13A3N).

Chapter Seven

An Integrated Approach to Achievement

7.1 Introduction

The main aim of this chapter is to introduce the method of Integrated Approach to Achievement (IAA). It is a new method that I have devised because no previous method can *very precisely* measure correctness, distinguish between levels of students, take error severity into account and study errors and correctness at the same time. This chapter introduces this new method and tests it by applying it (together with its measures) to a sample of the present study data and comparing the results of two groups. Accordingly, the questions answered in this chapter are (1) What is IAA? (2) How does it work? (3) Can it be successfully applied to the data of this study? (4) How are the results of two groups compared to each other after the application of this method? and (5) How does this method differ from the other two methods of correctness analysis that are applied in chapters five and six.

This chapter provides a detailed description of IAA. Following the description of the method, the chapter presents a clear and thorough explanation of the application of IAA to three essays as examples from the set of data collected for the present study. The choice of these examples has not been arbitrary but based on a number of points stated in the section devoted to this topic. The chapter includes, as well, a section that demonstrates the results of applying this method to the same 80 essays that have been analyzed using the method of T-unit and clause-based correctness analysis and various-units-based correctness analysis in chapter five and six respectively. The process of engaging another teacher in the application of this method is also discussed in section 7.4 of the chapter. The method is compared to the methods of T-unit and clause-based correctness analysis and various-units-based correctness analysis in section 7.5.2. The conclusions of the chapter are given in section 7.6.

7.2 A description of IAA

IAA works principally on a four-category T-unit basis: (1) the completely correct T-units (CCT), (2) the not completely correct T-units (NCCT) i.e. T-units with simple errors, (3) the incomprehensible or unanalyzable T-units (UT) and (4) the not completely correct T-units due to sentence boundary missing and confusion punctuation errors (NCCT-P). The method includes a category other than T-units, that is fragments. If the stretch of language is a

fragment not a T-unit. (e.g. a noun phrase or a dependent clause followed by a full stop), it is labeled as a fragment, which could be correct or contain errors. This method is both meaning- and syntax-based. It is meaning-based because it takes into account, to an extent, the (in)comprehensibility of the T-unit, and it is syntax-based because it takes into account both erroneousness and correctness. Not only does it identify the unit that should be erroneous or error-free but goes further than this by identifying the type and number of errors in these units. The method even analyzes the category that is unanalyzable in terms of the various-units-based correctness analysis into correct and/or 'almost' correct units. It is worth mentioning that it is the nature of the data that required the postulation and application of such a comprehensive method that is syntax-, meaning- and context-based. Also, I agree with Gaies (1980:55), as he casts doubt on "whether an index … divorced from *considerations of appropriateness* and *stylistic effectiveness* [my emphasis] can be a valid measure of overall language proficiency".

The following is a step-by-step explanation of how the method of IAA works:

1. Read the stretch of language well;
2. If it is a fragment, mark it as a fragment that is either, internally 'well-structured' - correct -(FC), or incorrect (FI);
3. If it is a complete T-unit mark it as either (1) a completely correct T-unit (CCT) if it is fully free from errors, (2) a not completely correct T-unit (NCCT) if it has a few simple errors (not more than 3 minor errors and/or not more than one major error) and carry out error analysis, (3) an unanalyzable T-unit (UT) if it contains more than 3 minor errors and/or more than one major error, and then break the identified UT down into smaller units (structures) that are either completely correct (CCS) or not completely correct structure (NCCS), with *only one error* and identify errors in the NCCSes.
4. If a T-unit is only incorrect because of punctuation confusion or a punctuation missing error at sentence boundary, label it as NCCT-P. Sometimes the same unit is identified as both NCCT-P and NCCT or UT when it contains errors other than the sentence boundary punctuation error (An algorithmic description of the method is provided in figure 7.1).

All errors dealt with in chapter four and presented in appendix G (the Université Catholique de Louvain's taxonomy of errors) are considered in the judgment of correctness except for spelling errors (only morphological spelling errors are considered and also if a spelling error led to the production of another word instead of the word meant to be in that position, see section 7.3 example 7.6 point 16). The Université Catholique de Louvain's taxonomy of errors is also used for describing and coding the errors. As for breaking the UTs into CCSes and NCCSes, the unit can be either identified as a CCS if it contains no errors and NCCS if it contains *only one minor* error. If the NCCS contains more than one minor error, it will be either further divided into smaller CCSes and NCCSes if possible or excluded from the analysis altogether if analyzing it further is not possible. The CCSes and NCCSes can be phrases of different types, or clauses of different types or two or three word units like *day to day*, *as simple as* etc. Units other than these are not considered. All errors are regarded minor except for the following:

- All lexical errors;
- Style or sentence errors when the sentence is unclear or very ill-formed.

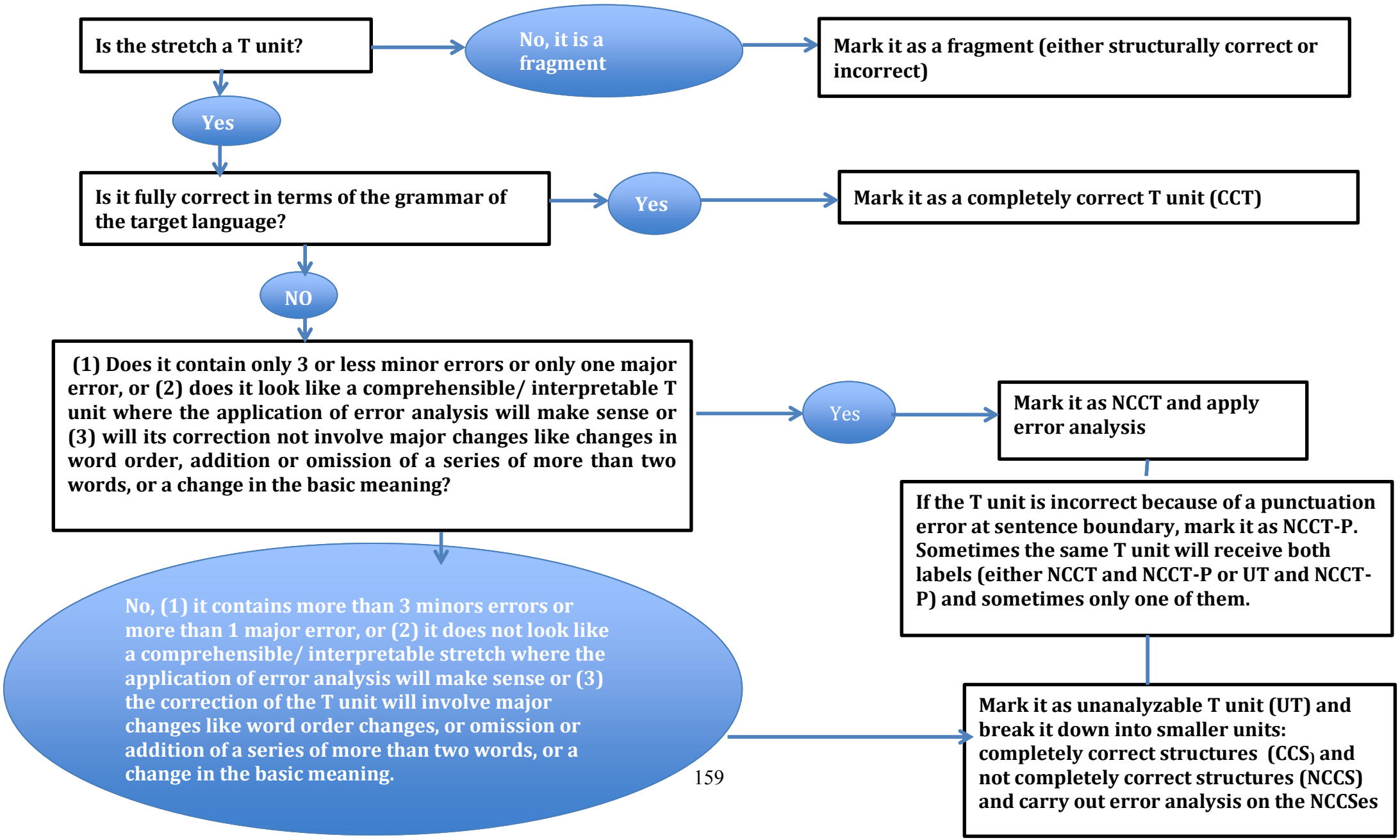
It is important to mention here that unlike in the other chapters (chapter four, five and six), errors in punctuation within sentences are considered with this method of analysis (see example 7.1). However, these errors do not play a role i.e. are not included within the number of errors when deciding whether the T-unit is NCCT or UT.

e.g. 7.1 S(140)

putting myself together and walking through the crowd, {QR} was the first step of getting rid of shyness.

In this example, there is an error of punctuation (a punctuation redundant error or a redundant comma, as underlined).

Figure 7.1 an algorithmic figure outlining IAA



To conclude, this method can be described in the following points:

1. It is meaning- and syntax-based;
2. It incorporates both correctness and error analysis;
3. It is a flexible method in terms of taking account of different units of measurement like making use of T-units (on the first level of analysis), and phrases, clauses etc. (on further levels of analysis);
4. It takes error severity into account;
5. It can provide a good method for writing pedagogy by presenting these CCTs NCCTes, CCSes, and NCCSes to the students to identify as either correct or incorrect, correct the incorrect , and then put them all in a wider context i. a whole paragraph in the case of CCTs and NCCTes and a whole sentence in the case of CCSes and NCCSes.

7.3 Three examples

In order to demonstrate how the method of IAA works, it is applied to three examples from the data (see appendices M, N, and O) and this section is devoted to the details of this application. The choice of these specific examples is based on a number of points. Student 5 is selected because when a sample of the data was analyzed using error analysis, it emerged that this student's essay contained six sentences that are too ill-formed to be analyzed with error analysis. Likewise, it contained seven unclear sentences, which either could not be understood from the context or their comprehension would be highly dependent on the reader's intuition (see appendix H). In addition to this, to show how this method works with the very low level writing, I have presented the analysis of essay 18, as it could be labeled as one of the very low level essays in the sample of the data. Concerning essay number 70, it has been chosen because analyzing it with the method of T-unit and clause-based correctness analysis and various-units-based correctness analysis, I came to the conclusion that this essay contains many of the T-units that are simply labeled as incorrect because of only one simple error. Analyzing it using the current method will, therefore, make it clear if this conclusion can be proved right.

The following symbols designate a number of units

// // = T-unit

( ) = the main categories either CCT, NCCT, NCCT-P, UT, FI, FC

[ ] = the smaller units (i.e. CCS and NCCS) that resulted from the analysis of UTs

{ }= errors and their corrections

1. Example: 7.2 (student number 5, see appendix M). It is important to highlight that not all errors in these essays are tagged but only the ones within NCCTs and NCCSes. The underlined parts are either CCSes or NCCSes. As for error tagging and correction, first the correction is added and then the tags within the same brackets.

1. First week at college (*FI*)
2. // It {*is, WMS*} different for all students // (NCCT*).*
3. //I thought because student {do, GVAUX} not know [NCCS] more about college [CCS], life in college[CCS] very different [CCS] for student{them, GPP}[NCCS] // (UT).
4. //So, student {s, GNN} saw another life not like life in school because, {0, QR} life in school was very routin, {0, QR} and easy. // (NCCT).
5. //Studen when came {to, XVPR} college [NCCS] they not saw that {the, GDD} life they saw in school [NCCS]//(UT).
6. //Maybe student {-s, GNN} in school [NCCS] had life very nice [CCS], and interesting //(UT).
7. //Students always in school had {have, GVT} many friends or best {good, GADJS} teachers //(NCCT) (NCCT-P),
8. //students in school had {undertook, LS }{a, GA} very simple study //(NCCT-P) (NCCT)
9. //always teacher {s, GNN} {and, LCC} their family {ies, GNN}help them not like in college, //(NCCT) (NCCT-P)
10. //maybe many students[CCS] when they came{to, XVPR} colleg[NCCS] not stayed at home[CCS]//(UT).
11. //They farther nom//(UT).
12. My first week at college (FC)
13. //that week was very difficult for me [CCS] because, when I came {to,XVPR) college [NCCS] I didn't saw that {the, GDD} people I saw in school [NCCS], and not saw that {the, GDD} friends I had in school [NCCS]//(UT).
14. // when I came{to,XVPR} college [NCCS] {the, GA} first day[NCCS] when I came [CCS] my feel is very sad [CCS], sorrowful//(UT).
15. Because for the first time I far my family and especially my parents my best friends (FI).

16. //when school always I thought with {about, XVPR} my best friends [NCCS]//(UT).
17. //We said {thought, LS}, {0, QR} we would go to {the, GA} same college//(NCCT)
18. //and that tim I didn't think we would seprat {from, XVPR} each other.//(NCCT)
19. //I came {to, XVPR} {this, GDD} college.// (NCCT)
20. //My best friend went {to, XVPR} another college // (NCCT-P) (NCCT)
21. //that's why I didn't saw {see, GVM} {my, GDO) best friend// (NCCT)
22. //and my teacher{s, GNN} {were, WMS} always helpful {to, XADJCO) for me//(NCCT).
23. //I didn't know life in college [CCS] and didn't know the teacher{s, GNN} [NCCS] and didn't know {the, GA} students [NCCS],// (NCCT-P), (UT)
24. //for the first week [CCS] that {those, GDD} days [NCCS] I didn't have friends [CCS] because, all students [CCS] didn't conects each others,// (NCCT-P), (UT)
25. //that {those, GDD} days {I, GPP} always thought {about, XVPR) my family, parents, brothers, and sisters. //(NCCT)
26. All days for the first week. (FI)
27. //I called my father, {0, QR} and mother// (NCCT) (NCCT-P)
28. //I said life in college was very difficult [CCS] and my feel very bad [CCS] and asked how could stay in college [CCS]// (UT)
29. but, when I called my brother. (FC)
30. //My brother always advised me. // (CCT)
31. //He said {to, XVPR} me, "college was {is, GVT} very good {“ , QM}// (NCCT) (NCCT-P)
32. //life in college[CCS] students {are, WMS} free [NCCS] more than {in, WMS} school [NCCS]. //(UT)
33. //All that {those, GDD} days when I went {to, XVPR, the, GA} upart ment {, QM}I cried for my family.// NCCT
34. //I missed them. //(CCT)
35. //Another day in that {the, GDD} first week {, QM} when I came {to, XVPR} class {, QM} I saw some teacher {s, GNN} // (NCCT-P), (NCCT)
36. //I didn't know them (CCS) because when I school I know all teacher, and always contected them [CCS] because some of that teacher lived in our village [CCS] and I love{GVT} them so much [NCCS]// (UT)
37. //My name (CCS) came in depart English// (UT)

38. that for the first time when came class. (FI)
39. //I saw all {the, GA} teachers spoken { speaking, GVV} English were {0, WRS} very fast.// (NCCT)
40. //I didn't understand one words {word, GNN} {from, GPR} them//(NCCT) (NCCT-P)
41. //I thought college was very difficult.// (CCT)
42. //I couldn't stay in my feel was very bad [CCS], // (NCCT-P), (UT)
43. //suddenly I thought [CCS] failed in college [CCS]// (NCCT-P) , (UT)
44. //I was very worry {worried, GWC} about all subjects.//(NCCT)
45. //I could {not, NP} studied {study, GVM} them because I couldn't understand them.// (NCCT)
46. //Day by Day for me {it, GPP} become {became, GVT} difficult in that {the, GDD} first week . //(NCCT)
47. //I couldn't ask {the, GA} teachers about subjects//(NCCT-P), (NCCT)
48. //I {was, WMS} always worried////(NCCT-P), (NCCT)
49. //always in that time I felt alone {lonely, LS}// (NCCT-P), (NCCT)
50. //nobody {was, WMS} near me // (NCCT-P), (NCCT)
51. //nobody {was, WMS} with me because I didn't know nobody {anybody, GPI} in the college.// (NCCT)
52. //I saw some students together. //(CCT)
53. //My parents always called me and advised me [CCS] for that feel bad. //(UT)
54. //When I went {to, XVPR) school {, QM} I lived with my parents {, QM} with my brothers, and sisters.//(NCCT)
55. //When I came home {, QM} all {the, GA}family helped me// (NCCT-P) , (NCCT)
56. //{the, GA}family helped me for {with, XVPR} subjects, food, washing, {and, LCC} cleaning,// (NCCT)
57. //and all worke {worked, GVM} for me, //(NCCT)
58. //but I came {to, XVPR} college [NCCS] I far {from, XADJPR} home [NCCS]//(NCCT-P) , (UT)
59. //nobody had help {helped, GVT} me //(NCCT-P), (NCCT)
60. //nobody work me because I far {from, XADJPR} home, [NCCS]// //(NCCT-P) , (UT)
61. //I live {in, XVPR} Akre not Duhok.// (NCCT)

62. // I didn't forget all that {those, GDD} days for {of , GPR} the first week.//(NCCT)
63. //I couldn't {cannot, GVT} forget them forever.// (NCCT)

Table 7.1 shows the results of the application of the method of IAA to essay number 5. It shows the number of CCTs, NCCTs, UTs, NCCT-Ps, CCSes, NCCSes, FCs, FIs together with their ratios. It also shows the number of errors in the NCCTs.

Table 7.1 the results of applying IAA to essay 5

Subject Code	CCT	NCCT-P	NCCT	No. of errors	UT	CCS	NCCS	FC	FI	CCT/T	NCCT-P/T	NCCT/T	UT/T
5	4	20	34	70	19	25	21	2	4	0.07	0.34	0.59	0.33

As table 7.1 shows, essay 5 contains only four completely correct T-units, 20 T-units that have been categorized as incorrect due to errors in sentence boundary punctuation, 34 NCCTs (T-units with simple errors) containing 70 errors, and 6 fragments (FI, FC). What is mostly noteworthy here is the number of the T-units that are labeled unanalyzable i.e. wrongly structured to the extent that they could not be analyzed in terms of error analysis and hence identified as (UT). However, as obvious from the table, these 19 UTs comprised a good number of (CCSes) and (NCCSes), which might be a good indicator that this student has a good repertoire of small units but does not find it easy to put these small units into bigger ones. To put it another way, the student does not have enough knowledge about the syntactic relations with which two- or three - word units are joined into sentences and sentences into a whole discourse. Consider the following examples:

e.g. 7.3

//maybe many students [*CCS*] when they came {to, XVPR} colleg NCCS] not stayed at home[CCS]//(UT).

e.g. 7.4

//I said life in college was very difficult [CCS] and my feel very bad [CCS] and asked how could stay in college [CCS]// (UT).

These two examples confirm what is mentioned above. The student seems to know that *many* should be followed by a countable plural noun and the noun *home* is always preceded by the preposition *at*. It is, however, evident that the student does not have the competence of the

syntactic rule that in English it is not correct to use two subjects for one clause i.e. *students* and *they*. So, instead of writing *maybe when many students came,* the student chose to write *maybe many students when they came.* Also, the student could not relate the verb *came* with the noun *college* by the preposition *to* committing another error in connecting the clause *when they came* with its complement *college.* In the second example, also, the three units *life in college was very difficult, very bad,* and *in college* have been slotted together in an ill-formed sentence.

As one can notice from the table 7.1, the student has scored very low on the measure CCT/T (0.07) but higher on both measures NCCT/T and UT/T (0.59, 0.33 respectively). This points to the fact that this students' written performance can be defined as a mixture of NCCTs and UTs. Consequently, it is difficult to apply error analysis without dividing the language into analyzable and unanalyzable units. The UTs are, nevertheless, further divided into correct and 'almost' correct units making it easy to compare this student's essay to another essay containing a similar number of UTs.

Now we turn to essay 184, which is considered to be one of the examples of very low-level essays. The results of the application of IAA to this essay are summarized in table 7.2.

2. Example: 7.5 (student number 184, see appendix N)

1. //Healthy is the most happfuly things to make [CCS] a best life.// (UT)
2. //Good healthy gives us the greate key [CCS] of clearly futures,// (NCCT-P) (UT)
3. //so human's depending on his/her environment [CCS] to be a truth of life [CCS].//(UT)
4. //As well as [CCS] life will be change by the places [CCS] as human's {humans', GNC} choice [NCCS],// (UT)
5. //but most of {0, WRS}people say [NCCS]; healthy is a part of [CCS] environments. //(UT)
6. //I do agree that the countryside life is much healthier than the city life [CCS] because, countryside has much way to healthy like, environment place and crowded by other ways [CCS].// (UT)
7. //Environments are so influnce of people's healthy {health, GWC} [NCCS] // (NCCT-P), (UT)

8. //so people by the environments can get [CCS] best healthy that is to be the way [CCS] of your wantted to choice. //(UT)
9. //On the other-ways environment helps you [CCS] to getting frish air [CCS] with {an, GA} active body [NCCS] also to be the kye of success in your plans [CCS] or will be the dreams [CCS].// (UT)
10. //Croweds are the other ways [CCS] of healthy that is do in the country side [CCS], because as we see [CCS] crowed in citys are most than the country side [CCS]. //(UT)
11. //As well as [CCS], city croweded has influenced healthy and your plans[CCS] because, your plans[CCS] related your health [CCS]. // (UT)
12. //most people [CCS] thinks leave the city [CCS] by the croweded ways//(UT) (NCCT-P)
13. //also government's tried for keeping it like, factories, cars, also make some other buildings [CCS] in outsides// (NCCT-P), (UT)
14. //but the main reseans {reason, GNN} of {a, GA} crowed is {are, GVN} people.//(NCCT)
15. //Tonicity is the most greatfuly of things to make frish life.//(UT)
16. the best way of tonicity to create the big kye of accually fiuture (FI)
17. //Also people depending on his/her environments to be truth of life [CCS].// (UT)
18. //Tonicity changed by the place [CCS] that is the frish air [CCS] like which place has more tries {trees, FS} [NCCS].//(UT)
19. //Also accepts countryside life relate with your healthier as they have better frish air and water [CCS].//(UT)

Table 7.2 the results of applying IAA to essay 184

Subject Code	CCT	NCCT-P	NCCT	No. of errors	UT	CCS	NCCS	FC	FI	CCT/T	NCCT-P/T	NCCT/T	UT/T
184	0	4	1	3	17	30	6	0	1	0	0.22	0.05	0.94

This student's essay seems to be of a lower level than essay 5 because, on one hand, the ratio UT/T is higher than that of essay 5 (0.94 compared to 0.33) and, on the other hand, the ratio NCCT/T is much lower (0.05 compared to 0.59). Furthermore, the ratio of CCT/T is also higher in the case of essay 5 than in the case of essay 184. It is obvious from this student's

score on UT/T (0.94) that his/her essay is merely a composition of stretches of language that are incomprehensible. Adding to this, unlike the previous student who wrote essay 5, this student looks to have very little knowledge of composing small units like phrases, multiple word units, clauses etc. This is because his/her UTs seem to be less 'structured' than the UTs of the student who wrote essay 5. What demonstrates this is that in the structure of a total of 17 UTs only 36 CCSes and NCCSes are identified, while in the case of essay number 5, 46 CCSes and NCCSes from the structure of 19 UTs have been extracted.

Now we turn to the description of the detailed application of IAA to essay number 70. The results are summarized in table 7.3

3. Example: 7.6 (student number 70, see appendix O)

1. //College is a place where people get educated, meet new friends and share information together.// CCT)
2. //{In, GPR} My first week at College, I felt something new in my personal life because I could meet different people from various areas// (NCCT)
3. //and I felt really happy about it,// (NCCT-P)
4. //at the same time I was somehow scared {of, XADJPR} how to deal with them due to their personalities and behaviours. //(NCCT)
5. //I could do great things in my first week of College like preparing presentations// (CCT)
6. //and Collaborating {with, XVPR} the other friends was something exciting.// (NCCT)
7. //I tried to choose some people at College and make them close-friends according to my own personality, //(NCCT-P)
8. //<u>everyone couldn't not be a friend</u> [CCS] of <u>my occupation</u> [CCS] //(NCCT-P), (UT)
9. //{the, GA} first week at College, I was missing (missed, GVA} my family as well, {0, QR} because it was my first time to get far away from my mother, father and brothers,// (NCCT-P), (NCCT)
10. //I realized how tough their job was,// (NCCT-P)
11. //so I tried to be punctual and serious {in, XADJPR} to my studying because my father over-worked to get money for the sake of me {my sake, LP} and I highly appreciated his fatigue. // (NCCT)

12. //When I first came to College {,- QM} my hobby was that to be a brilliant translator and make my dream come true.// (NCCT)
13. //My friends were very helpful with me at {in, GPR} the first days at {of GPR} College// (NCCT)
14. //and I didn't even feel that I am forigen among then {them, FS} because whenever I had difficulties with some topics, they would be very kind and help me without taking {making, LS} any excuses.// (NCCT)
15. //In my first days at {the, GA} aPartment we used to have much fun and work together in every single job, // (NCCT-P), (NCCT)
16. //we tried to cock {cook, FS} some food and learn from each other,// (NCCT-P), (NCCT)
17. //actually it was really interesting because sometimes we burnt the food and some other times it was very delicious.// (CCT)
18. //My Partners of a {at the, GPR, GA} Partment were from different places,// (NCCT)
19. //and each one had his own dialect,// (NCCT-P)
20. //So at first their dialect {dialects, GNN} was not really clear for me, //(NCCT)
21. //but after being together for a long time {, QM} I could deal with their own dialects.//(NCCT)
22. //In conclusion, College to me was an impressive place that taught me how to deal with sensitive, serious and funny people// (CCT)
23. //and I learnt many things from my teachers, friends and even my personal mistakes, // (NCCT-P)
24. //eventually I realized how to live my life and make us {use, FS} of it.// (NCCT)
25. //Although meeting various friends is not a matter of exploitation or having fun all the time, but {0, LCC} it is a process of cooperation and helping each other in all situations// (NCCT)

Table 7.3 the results of applying IAA to essay 70

Subject code	CCT	NCCT-P	NCCT	No. of errors	UT	CCS	NCCS	FC	FI	CCT/T	NCCT-P/T	NCCT/T	UT/T
70	4	9	15	21	1	2	0	0	0	0.16	0.36	0.6	0.04

This table (7.3) provides a clear justification that the conclusion explained above in the introduction of this section concerning essay 70 is right. Notice that most of the T-units in the essay are 'almost' correct. In other words, 15 T-units out of 25 T-units have been identified as

NCCTs. This is in addition to the fact that these have been marked as incorrect because of 21 errors (with an average of 1.4 of error/T-unit) i.e. the majority of these NCCTs have turned out to be erroneous just because of one simple error. It is also obvious that these errors are very simple errors (e.g. spelling errors that led to the use of a different word like the use of *cock* for *cook* in T-unit 16 and *us* instead of *use* in T-unit 24, punctuation errors like the absence of comma in the T-unit 12, errors in prepositions like the lack of the preposition *of* in the T-unit 4 etc.). Investigating further, 9 out of 25 T-units are defined as incorrect merely due to punctuation errors (at sentence boundary)[1]. All these facts about the performance of this student have not been explored when analyzing his/her performance by only identifying the error-free T-units in chapter five (EFT/T =0.24). To state this more clearly, this student seems to have been very unfairly disadvantaged as his/her essay has not been accurately judged in terms of accuracy.

7.4 User engagement

As with the previous analyses and in a discussion with me as the first rater, 20% of the data analyzed for this chapter have been reanalyzed by another rater who is specialized in linguistics and hence is likely to use the research. The discussion took place after I explained all the process and rubrics of analysis to the second rater. The discussion revealed different points of agreement and disagreement, although the points of agreement tended to be more. This is because of the presence of clear criteria for dividing the texts into FC, FCI, CCTs, NCCTs, NCCT-Ps and UTs, CCSes and NCCSes. The points of disagreement included (1) clear cases that we soon reached an agreement on and (2) a number of other cases, which remained open to discussion.

The following are examples of the first case:

e.g. 7.7 S(4)

No one can live without facing saddest event or funniest event (1st T-unit) *we exersice both of them* (2nd T-unit).

I have considered the second T-unit CCT but the second rater has considered it as NCCT correcting it to *We experience both of them*.

e.g. 7.8 S(1)

As for us as children we went next to the water.

I have coded this as NCCT but the second rater considered it CCT.

e.g. 7.9 S(7)

At the End my advice to all the student who are coming to the college make sure that you have chose good department, so you will be happy at the end.

We both agreed that this is a UT but in regard to its analysis further to CCSes and NCCes, I have not coded *good department* as an NCCS which could only be corrected by adding the indefinite article *a*. The second rater noticed this and identified this as NCCS.

e.g. 7.10 S(6)

We should have a great and pure heart in order to help people out who is in a difficult situation, because I was in a hard and not good situation.

We both agreed that this a UT but I have missed the clause *who is in a difficult situation* as a NCCS. The second rater noticed this and regarded this as a NCCS.

The following are examples of the second case, which remained open to discussion:

e.g. 7.11 S(1)

My family gathered with the relatives and had prepared for this journey

I have regarded this as an NCCT because of the use of past perfect *had prepared for this journey*. The second rater coded this T-unit a CCT basing her argument on the justification that the use of past perfect in *had prepared for this journey* is correct because this sentence contains two events: the family getting together with the relatives and the preparation for the journey, and it is more justifiable to say that the preparation for the journey has happened before the gathering. However, this example depends on what the writer means by the sentence. If he/she means that the family prepared for the journey and then gathered with the relatives in the same journey then the second rater is right but if the writer means that the family gathered with the relatives to prepare for the journey, then my coding is correct.

e.g. 7.12 S(9)

on one hand the first week is a week of saddness because of the difficult and the different style of having lectures and many other things but on the other hand it is a week of happiness, first because of getting the college and second because of getting friends and many other things .

In this T-unit, I have not considered *the difficult and the different style of having lectures and many other things* as an error but the second rater considered the use of the word *style* as a lexical error arguing that it should be either *method* or *way*.

e.g. 7.13 S(4)
so most of their people are poor

I have identified this as CCT but the second rater considered it as NCCT because, as she declared, the sentence should be *so most of <u>its</u> people are poor* because *it* here refers to Syria which exists previously in the text: *and in such country like Syria the life wasnot going well.*

e.g. 7.14 S(14)
but <u>class by class</u> I became better

I have ranked this as CCT but the second rater coded it as NCCT justifying her choice of this code by the point that such a phrase like *class by class* cannot be produced by a native speaker (i.e. a native speaker can say *day by day*, *week by week* but this cannot be extended to the word *class*).

7.5 Results and discussion

To achieve consistency between the scoring methods used, this method has been applied to the same 80 essays that were analyzed in chapters five and six. The sample consists of the students of two schools, school of Arts and school of Basic education (40 students per each school). The 40 students of each school comprise two groups (20 third year students and 20 fourth year students). First, a general demonstration and analysis of the results are presented; then a comparison is made between this method, the method of T-units- and clause-based correctness analysis and the method of various-units-based correctness analysis; finally the performance of the third year students is compared to the performance of the fourth year students (both schools).

7.5.1 A general demonstration and discussion of the results

The following figures (7.2, 7.3, 7.4, 7.5) demonstrate the results.

Figure 7.2 shows the ratios of Completely Correct T-units/ total number of T-units (CCT/T), Not Completely Correct T-units due to punctuation errors/ total number of T-units (NCCT-

P/T), Not Completely Correct T-units due to errors other than punctuation/ total number of T-units (NCCT/T) and Unanalyzable T-units/ total number of T-units (UT/T) for 20 essays written by 20 third year students/ school of Arts. The horizontal axis shows the students' (or essays) codes; for example, 1A3N is essay number 1 written by a student from the school of Arts (A) in third year of study (3) and its type is narrative (N). The vertical axis represents the ratios stated above.

Figure 7.2 results of applying IAA to a sample of third year students, school of Arts

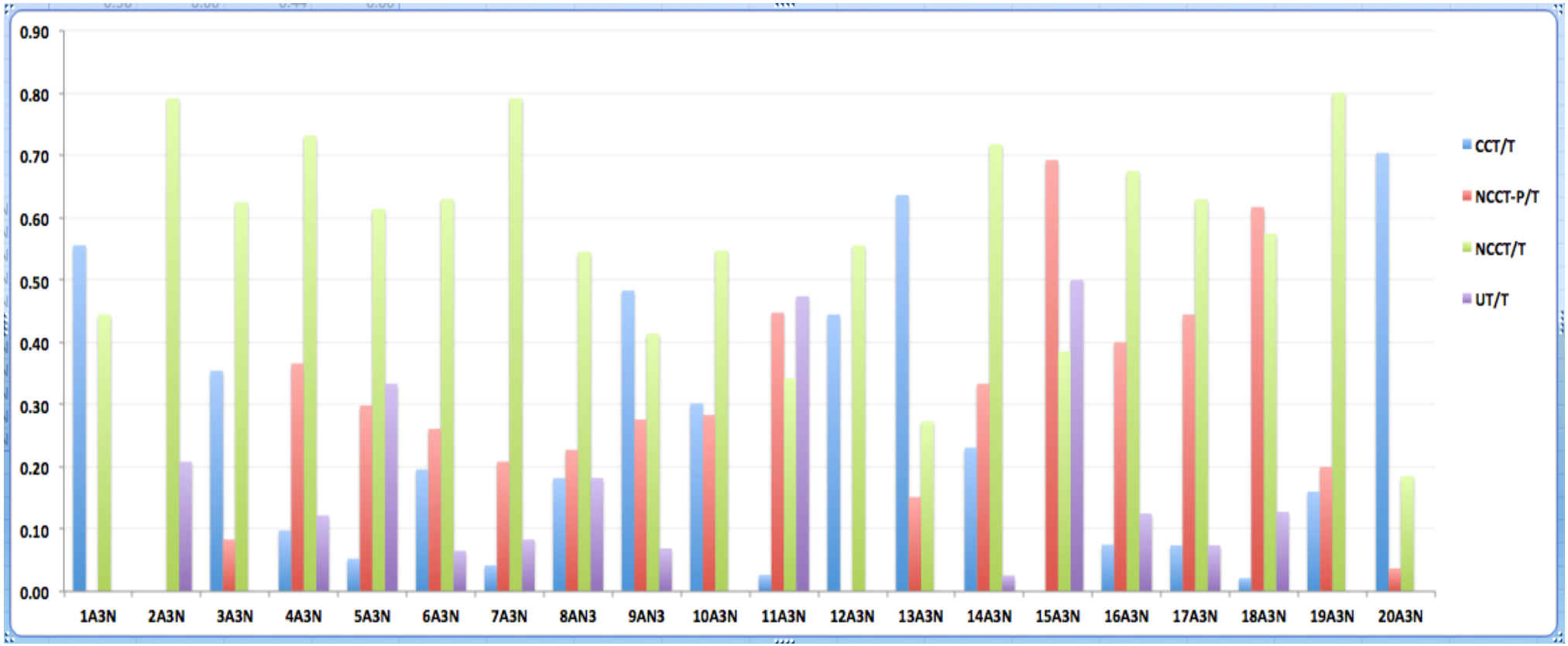


Figure 7.3 shows the ratios of Completely Correct T-units/ total number of T-units (CCT/T), Not Completely Correct T-units due to punctuation errors/ total number of T-units (NCCT-P/T), Not Completely Correct T-units due to errors other than punctuation / total number of T-units (NCCT/T) and Unanalyzable T-units/ total number of T-units (UT/T) for 20 essays written by 20 fourth year students/ school of Arts. (For more information about the abbreviations, see figure 7.2.

Figure 7.3 Results of applying IAA to a sample of fourth year students, school of Arts

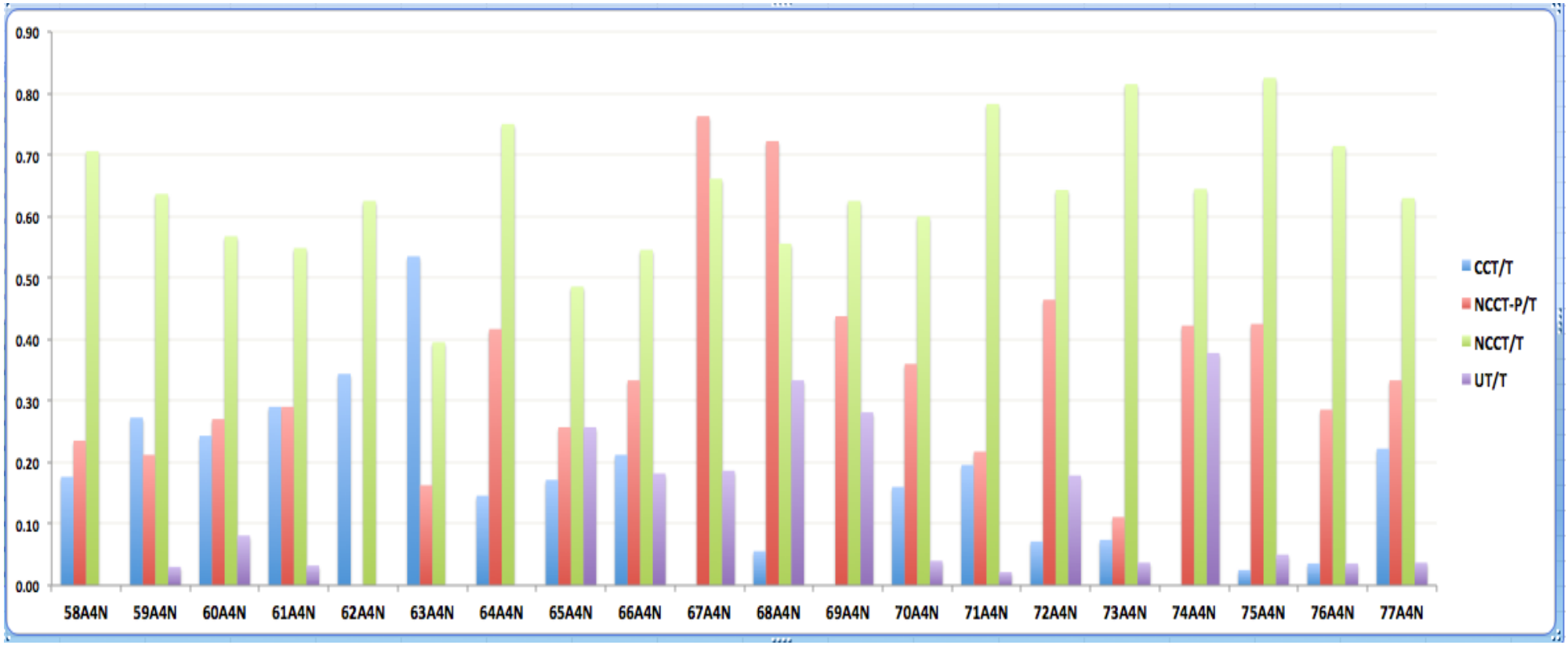


Figure 7.4 shows the ratios of Completely Correct T-units/ total number of T-units (CCT/T), Not Completely Correct T-units due to punctuation errors/ total number of T-units (NCCT-P/T), Not Completely Correct T-units due to errors other than punctuation / total number of T-units (NCCT/T) and Unanalyzable T-units/ total number of T-units (UT/T) for 20 essays written by 20 third year students/ school of Basic Education. The horizontal axis shows the students' (or essays) codes for example 113B3N is essay number 113 written by a student from the school of Basic Education (B) in the third year of study (3) and its type is narrative (N). The vertical axis represents the ratios stated above.

Figure 7.4 Results of applying IAA to a sample of third year students, school of Basic Education

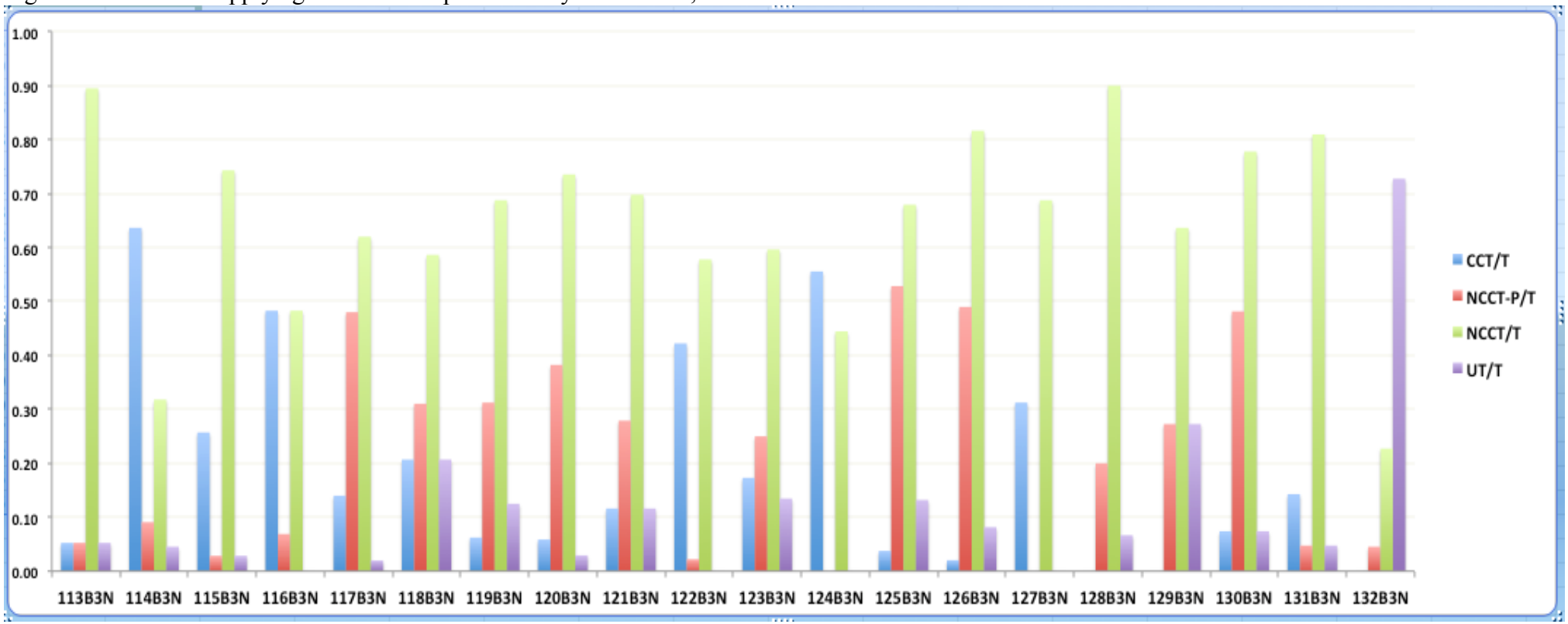
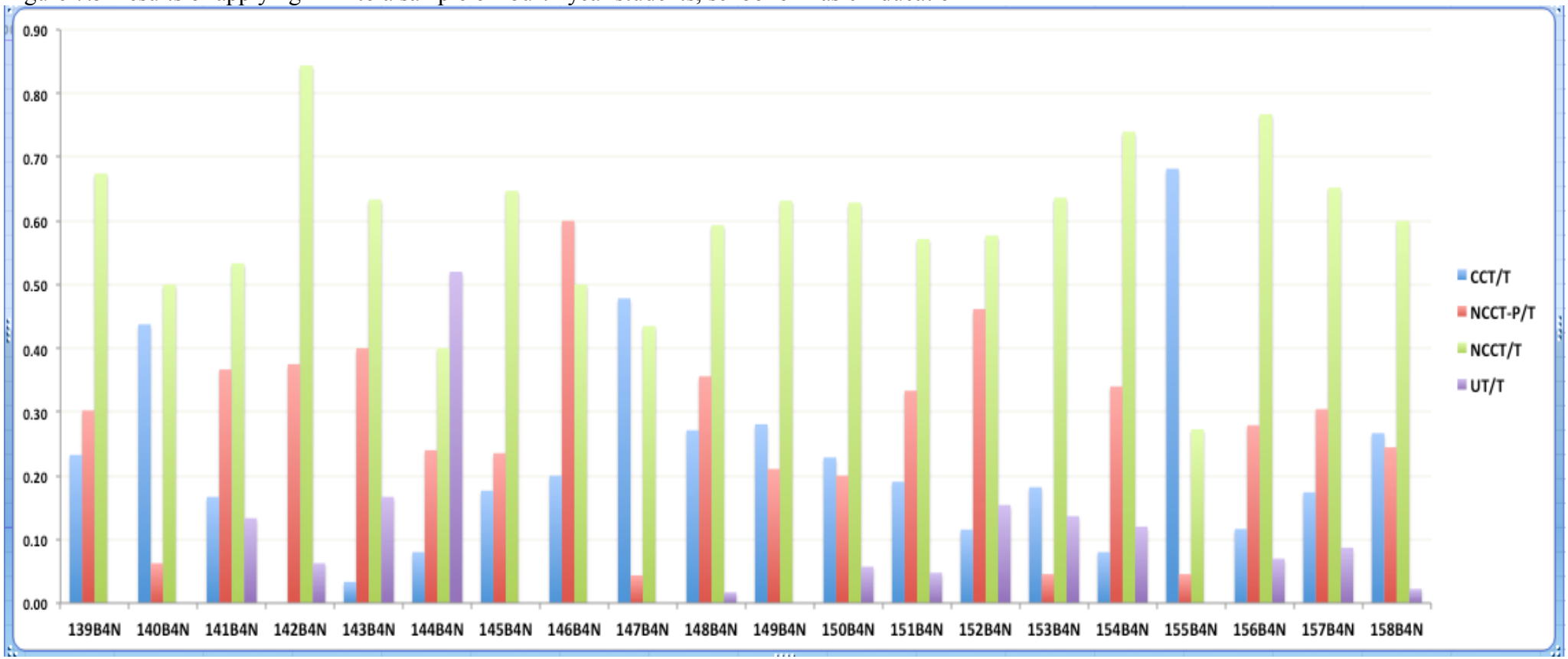


Figure 7.5 shows the ratios of Completely Correct T-units/ total number of T-units (CCT/T), Not Completely Correct T-units due to punctuation errors/ total number of T-units (NCCT-P/T), Not Completely Correct T-units due to errors other than punctuation / total number of T-units (NCCT/T) and Unanalyzable T-units/ total number of T-units (UT/T) for 20 essays written by 20 fourth year students /school of Basic Education. For more information about the abbreviations, see figure 7.4.

Figure 7.5 Results of applying IAA to a sample of fourth year students, school of Basic Education



As all the figures (7.2, 7.3, 7.4, 7.5) show, the ratio NCCT/T is the highest for nearly all students except for a very few. This indicates that most of the students' interlanguage has become erroneous due to a few number of minor errors (maximum three minor errors or one major error). This can be counted as an advantage of IAA since it does not make dichotomous distinctions between erroneous and error-free units where it is possible for one single minor error to render a whole otherwise well-formed T-unit incorrect. Also, according to this method, these NCCTs are different from UTs that are erroneous to the degree that they cannot be analyzed in terms of error analysis. Thus, this method takes account of error gravity as well as number of errors that make a T-unit incorrect. This method can also distinguish two NCCTs in terms of their severity of incorrectness through identifying the number and type of errors in these two NCCTs. For example, a student might produce more NCCTs but with a lower number of errors and errors with less severity. For this reason, other numbers in figures (7.6, 7.7, 7.8, 7.9) are given showing the average error per NCCT for each student. Even if the average number of error per NCCT is still not a sufficiently detailed indicator to assess the NCCTs, the researcher or assessor can look further at the types of errors in these NCCTs.

Now, for instance, comparing the NCCTs produced by the student 7A3N with those produced by the student 2A3N in figure 7.2 who has produced almost the same number of NCCTs (regardless of the other categories of T-units they have produced), one must look first at figure 7.6 to see the average number of errors per NCCT. If this still does not differentiate between the two, one can see the types of errors in the NCCTs. It is evident from figure 7.6 that student number 2A3N has a higher average of error per NCCT than student 7A3N (2.21 compared to 1.42, see appendix P). As for the type of errors committed by them within the NCCTs, student 7A3N has only 13 types of not very severe errors while the student 2A3N has produced 21 types of errors including multiple word missing. The case is the same with students 73A4N and 75A4N in figure 7.3. The performance of these two students is very similar especially their NCCTs but looking at figure 7.7, it becomes obvious that the NCCTs produced by student 75A4N contain fewer errors than those produced by student 73A4N. The same applies to students 131B3N and 126B3N in figure 7.4. The difference between these students cannot, however, be compared depending on the average error per NCCT because as it is clear from the figure 7.8, their mean error per NCCT is nearly equal. Accordingly, the types of errors they committed in their NCCTs were analyzed. Subject 131B3N has made only 17 types of errors that are less serious than the 21 types of errors committed by subject 126B3N. This is how one can make very fine distinctions between essays in terms of their accuracy.

Figure 7.6 the average number of errors per NCCTs produced by a sample of third year students/ school of Arts

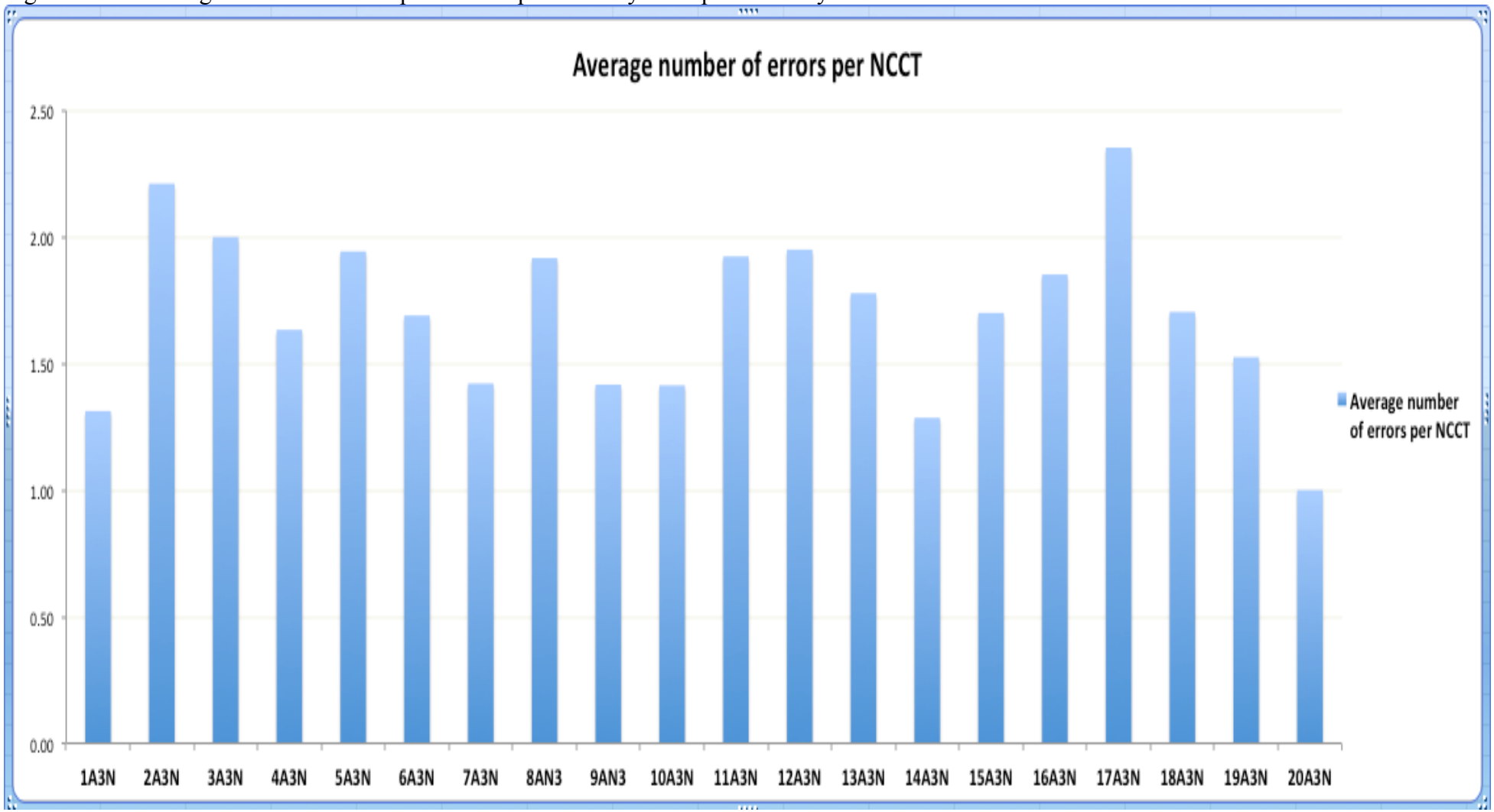


Figure 7.7 the average number of errors per NCCTs produced by a sample of fourth year students/ school of Arts

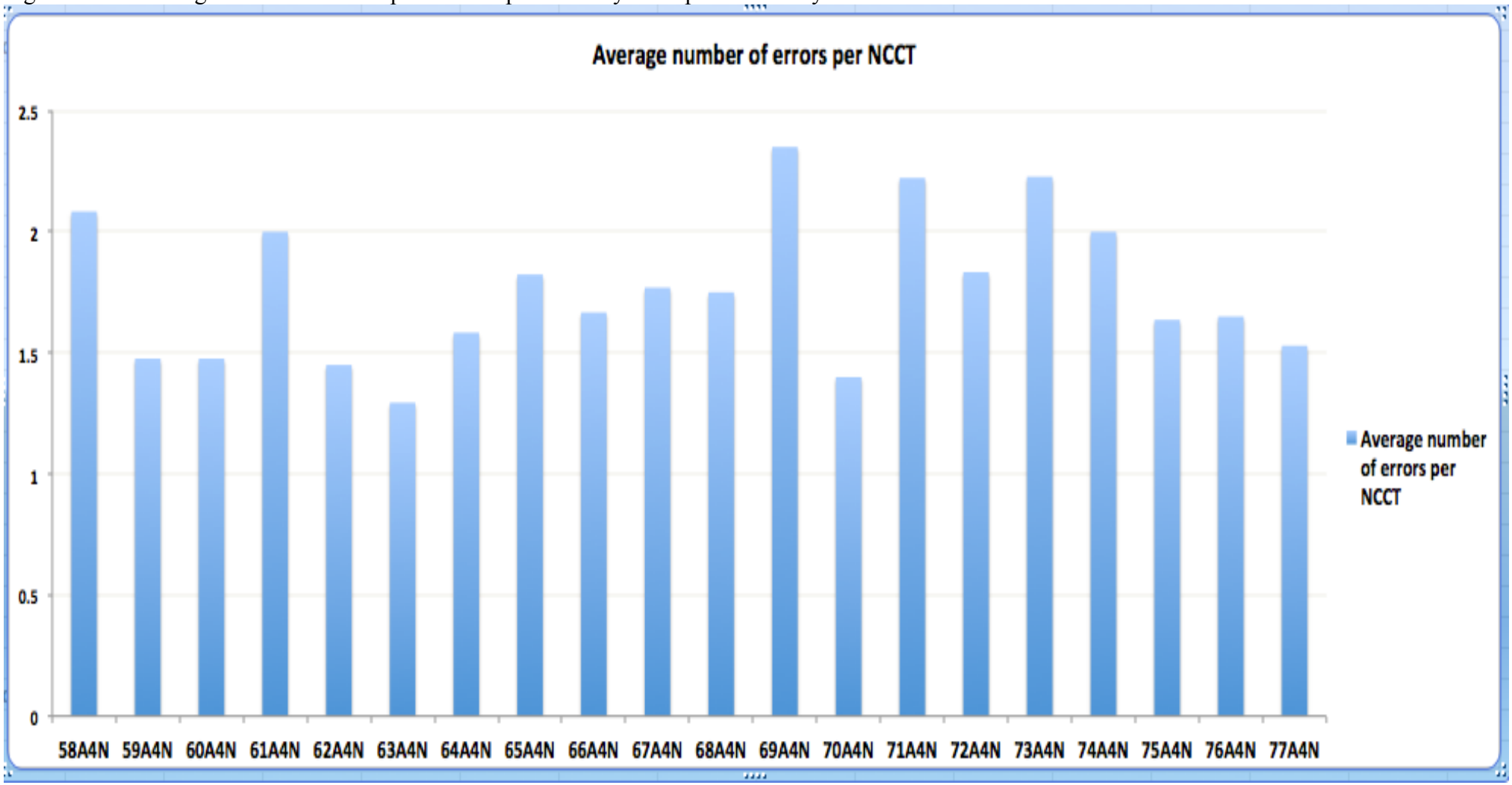


Figure 7.8 the average number of errors per NCCTs produced by a sample of third year students/ school of Basic Education

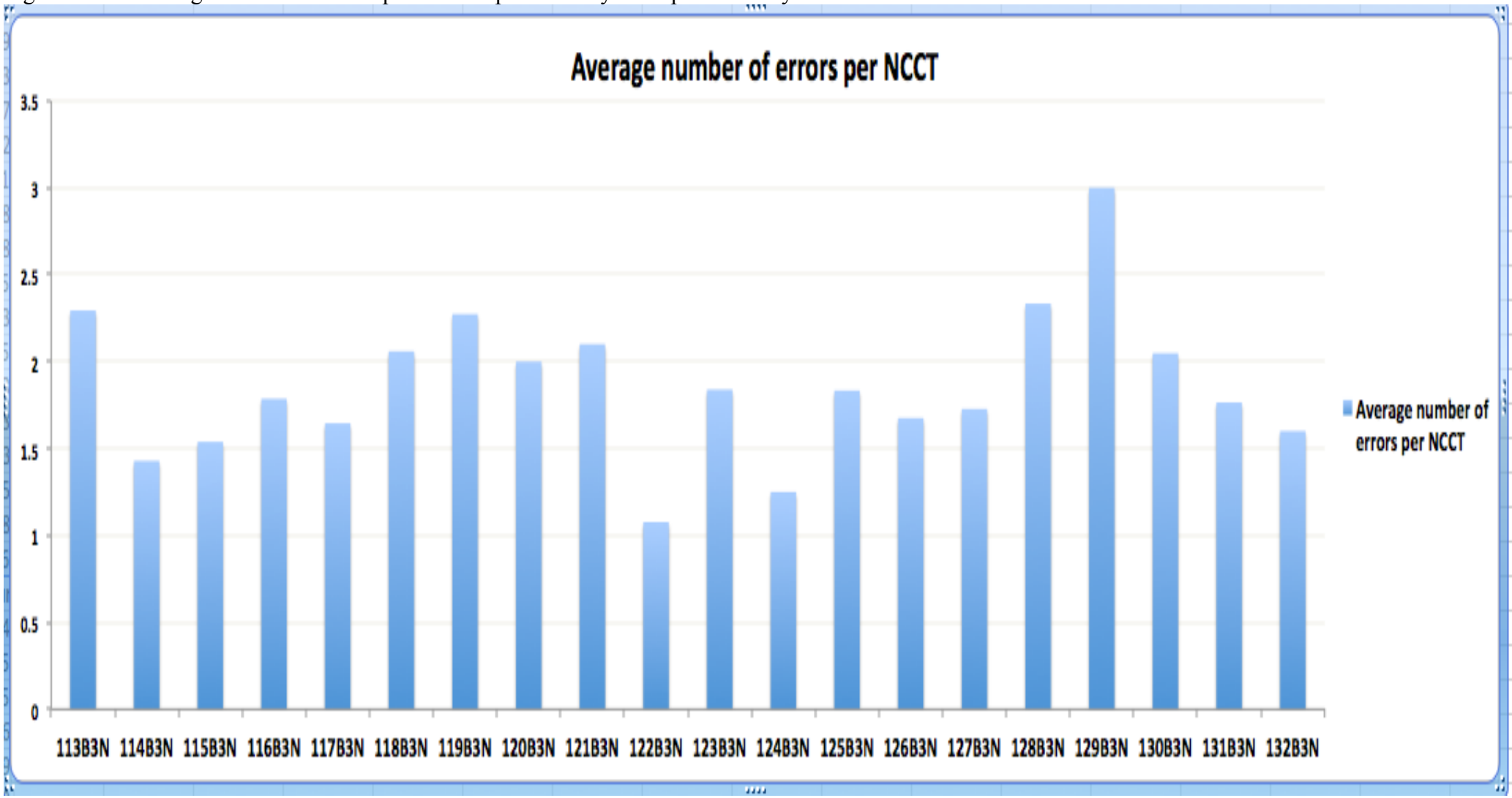
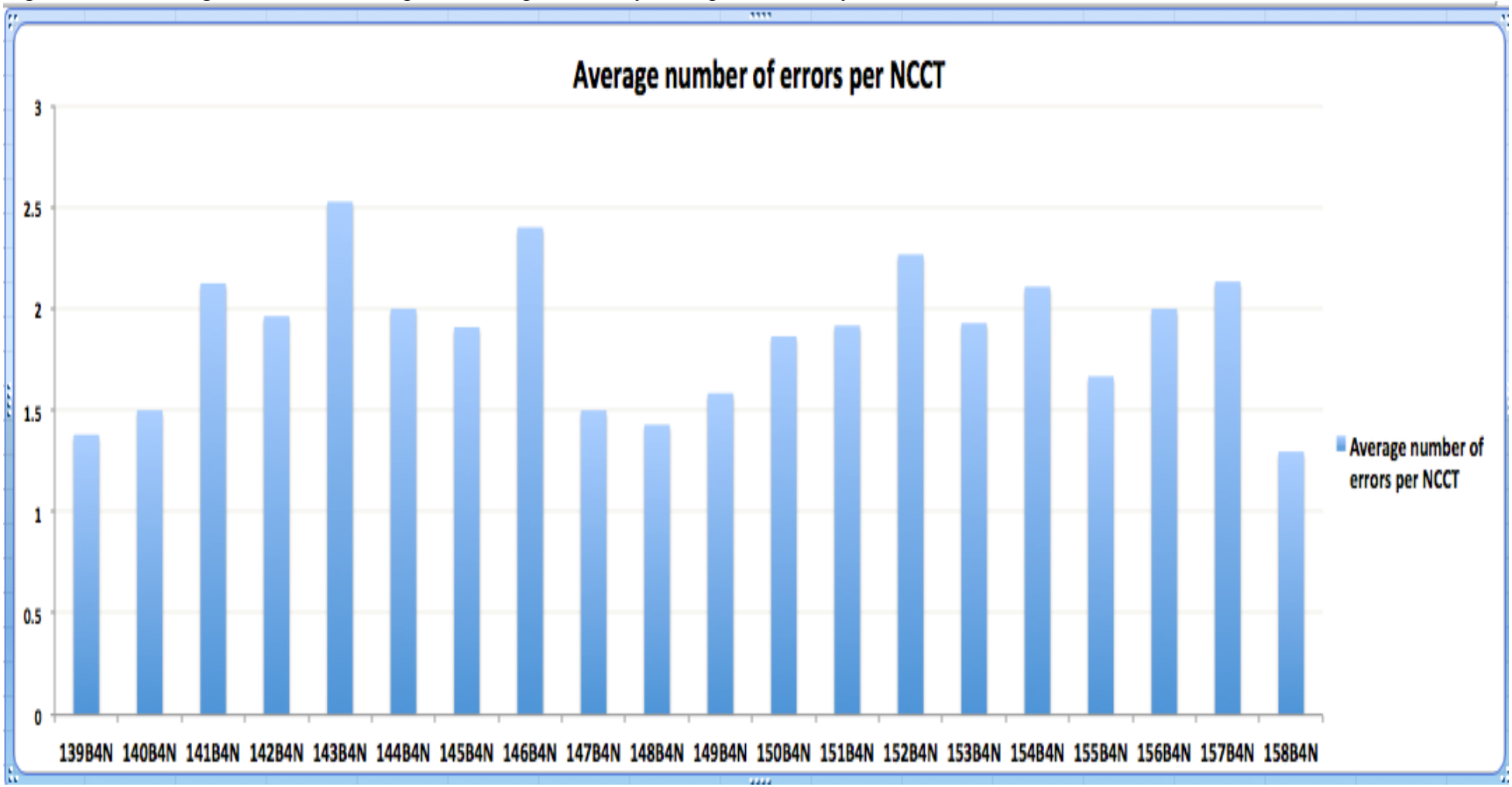


Figure 7.9 the average number of errors per NCCTs produced by a sample of fourth year students/ school of Basic Education



The unanalyzable T-units can also be compared to each other by counting the number of CCSes and NCCSes they comprise. The more CCSses and the more NCCSes in the UT, the more structured it is. Additionally, it is interesting to look at the average length of these CCSes and NCCSes because the longer they are, the more structured the UT is. If, for example, we compare the UTs of student 16 (5 CCSes and 4 NCCSes with an average length of 4.8 word per CCS and 3.5 word per NCCS, see table 7.4) to those produced by student 17 (only 4 CCSes with an average length of 2.25 word per CCS see table 7.4), we can evaluate student 16 as having produced UTs that are more structured than those produced by student 17. For this purpose, the mean lengths of CCSes and NCCSes of the UTs produced by one sample (third year students/ school of Arts) are listed in table 7.4 as an example.

Table 7.4 shows the number of CCSes and NCCSes in the UTs produced by the third year students/ school of Arts together with their ratio to the total number of the UTs.

Table 7.4 the ratios of CCS/UT and NCCS/UT with the average length of CCSes and NCCSes/third year students/school of Arts

Students codes	UT	CCS	NCCS	CCS/UT	NCCS/UT	Average length of CCS	Average length of NCCS
1A3N	0	0	0	0.00	0.00		
2A3N	5	14	3	2.80	0.60	3.92	5.33
3A3N	0	0	0	0.00	0.00		
4A3N	5	2	6	0.40	1.20	8	7.83
5A3N	19	25	20	1.32	1.05	3.2	3.55
6A3N	3	8	0	2.67	0.00	7.37	
7A3N	2	2	3	1.00	1.50	4	6
8AN3	4	15	3	3.75	0.75	3.8	4
9AN3	2	3	0	1.50	0.00	3	
10A3N	0	0	0	0.00	0.00		
11A3N	18	36	28	2.00	1.56	3.5	4
12A3N	0	0	0	0.00	0.00		
13A3N	0	0	0	0.00	0.00		
14A3N	1	4	0	4.00	0.00	4.25	
15A3N	13	17	5	1.31	0.38	4.76	5.6
16A3N	5	5	4	1.00	0.80	4.8	3.5
17A3N	2	4	0	2.00	0.00	2.25	
18A3N	6	14	5	2.33	0.83	4.28	6.6
19A3N	0	0	0	0.00	0.00		
20A3N	0	0	0	0.00	0.00		

As table 7.4 indicates, although students have produced a number of UTs, still these UTs are structured. Most of them were eligible to be divided further into CCSes and NCCSes, and

these are structures that are either phrases or clauses or two- or three word units. However, these CCSes and NCCSes are not so long except for those produced by the two subjects, 4A3N and 6A3N. If one notices the number of UTs produced by these students, one can clearly see that they have not produced a great number of UTs. This and the significant length of their CCSes and NCCSes could be a good indicator that, despite the fact that these students have committed errors, their errors are not severe to a degree that makes their UTs incomprehensible. In other words, their UTs are more comprehensible than the UTs that could only be divided into small CCSes and NCCses. This is because their language is broken down into smaller stretches due to errors but still these stretches are long enough to make sense. Consider the following examples of UTs:

e.g. 7.15 S(4)

At the end, people have to face difficulty inorder to behave or to remember that there is a God who created us who can get life to or die us.

e.g. 7.16 S(74)

but some events occure on ourself that we never do not think about it such as when we lost our friends because of accident cars or we should separated our family or relative.

The T-unit in example 7.15 is marked as UT (unanalyzable) according to the model used in this chapter because it has more than 3 errors. Nevertheless, it contains the long NCCS *people have to face difficulty inorder to behave or to remember that there is a God who created us*, where there is only one error (the word *difficulty*). On the other hand, the UT in example 7.16 is so unstructured that it could only be divided into the small stretches *some events*, *our friends,* and *our family,* only three phases with an average length of two words.

From a holistic view of these essays, I gained the impression that most of the students' writing was composed of UTs. However, the figures actually show a relatively low proportion of UTs. Recognition of calculations such as this might change teachers' view about assessing students' essays from only looking holistically at their work to reading it very carefully.

Also worth mentioning is the ratio of the T-units that have been considered incorrect because of sentence boundary punctuation (NCCT-P). It is clear from the figures (7.2, 7.3, 7.4, 7.5) that this is a serious problem in these students' written production of English. These T-units have all been regarded incorrect and excluded from analysis with the method of correctness analysis in chapter five.

7.5.2 A comparison between IAA, the T-unit and clause-based method of correctness analysis and the method of various-units-based correctness analysis

The points of difference between IAA and the other two methods of measuring accuracy namely, the T-units and clause-based correctness analysis and various-units-based correctness analysis can be summarized as follows.

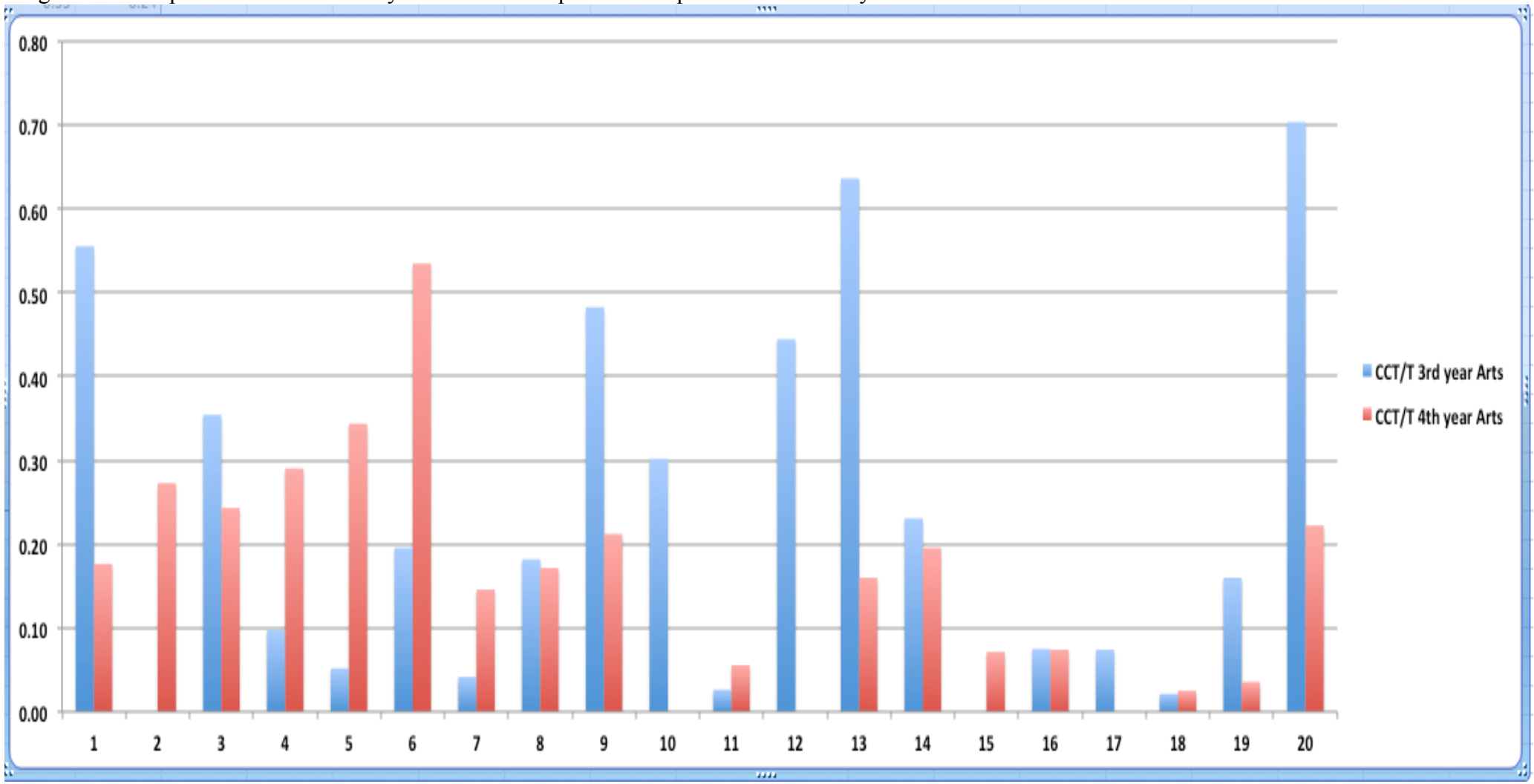
Compared to the correctness analysis used in chapter five (T-unit and clause-based correctness analysis), this method differs in three important points. The first one is measuring the students' achievement on different levels or a continuum scale rather than a dichotomous scale. In this way, the method is capable of providing a clearer view about the performance of the students rather than presenting their production as a mixture of correct and incorrect units. The second is that this method deals with errors in their own right as well as correct forms while the method of T-units and clause-based correctness analysis identifies only the correct forms and leaves both the incorrect T-units and the errors that cause this incorrectness unanalyzed. The third one is that this method accounts for both fragments and T-units that are incorrect because of sentence boundary punctuation errors.

This method also differs from that of various-units-based correctness analysis in two points. The first point is that this method takes one type of unit (T-unit) as a unit of analysis and later goes a level further by applying a method similar to various-units-based correctness analysis by breaking the unanalyzable T-units into completely correct sequences (CCS) and not completely correct sequences (NCCS). The second point of difference is that this method fills the gap of not considering both correct and incorrect units as it deals with correctness and incorrectness rather than dealing only with the correct units.

7.5.3 A comparison between the performance of the third year students and fourth year students of two schools /school of Arts and school of Basic Education

The following figure (7.10) shows a comparison between the performance of the third year students/ school of Arts and the performance of the fourth year students/ school of Arts as far as the ratio CCT/T is concerned. The horizontal axis represents the students of both years and the vertical axis represents the ratios the students have scored.

Figure 7.10 the performance of the 3rd year students compared to the performance of 4th year students /school of Arts in terms of the measure CCT/T



The following table (7.5) is a summary of the figure 7.10. It groups the students who have scored within the same ratio in one column.

Table 7.5 the ratio of CCT/T of year 3 compared to year 4/school of Arts
CCT/T: A summary of figure 7.10

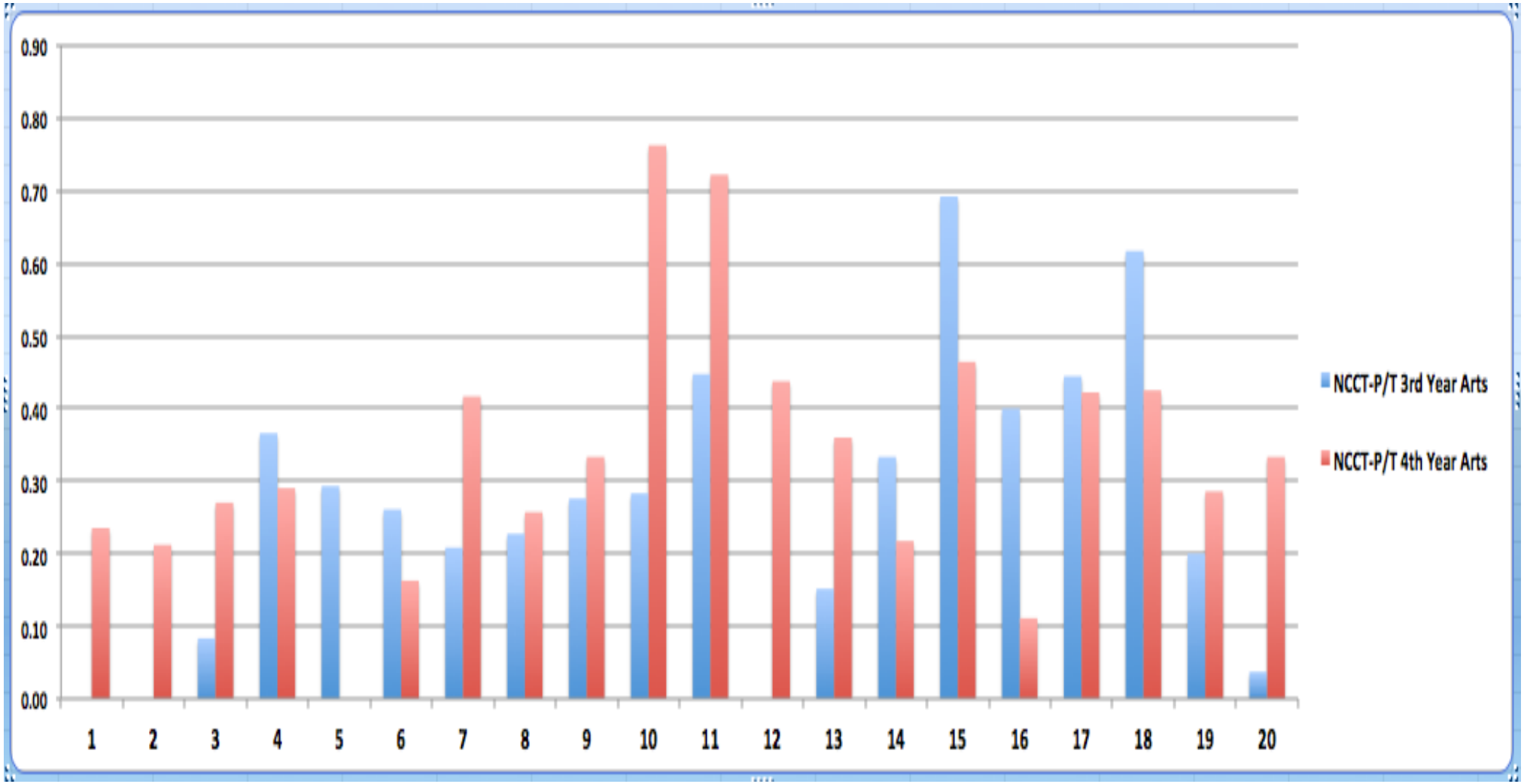
	0.00- 0.10	0.11-0.20	0.21-0.30	0.31-0.40	0.41-0.50	0.51-0.60	0.61-0.70	0.71-0.80	0.81-0.90	0.91-1
3rd year Arts	9	3	2	1	2	1	2	0	0	0
4th year Arts	8	5	5	1	0	1	0	0	0	0

As obvious from table 7.5 and figure 7.10, the fourth year students as a group seem to be more homogenous than the third year students as there is more variation in the levels of third year students than in the levels of fourth year students. Although both groups did not perform very well in terms of producing CCTs because the numbers of students is high on the low ranges (for example, the number of students in both groups who have scored between 0.00 and 0.10 is high in both cases), there seems to be a difference between third year students and fourth year students. More third year students produced a higher number of CCTs and this is obvious from the higher number of third year students than the fourth year students on the high scores and their lower number on the low scores (notice their numbers on the ranges 0.11-0.20 and 021-0.30 as an example of the low ranges and the ranges 0.41-0.50 and 0.61-0.70 as an example of the high ranges).

It is worth mentioning though that the two third year students who have scored between 0.61–0.70 are students number 13 and 20; as mentioned earlier in chapter five, these two students have learnt their English in a different context and most probably this is the reason why they scored higher than the others. Although there seems to be a difference between the students, the P value in the two-tailed T test I have conducted did not indicate that the difference is significant (P=0.22).

The following figure (7.11) shows a comparison between the performance of the third year students/school of Arts and the performance of the fourth year students/school of Arts as far as the ratio NCCT-P/T is concerned. The horizontal axis represents the students of both years and the vertical axis represents the ratios the students have scored.

Figure 7.11 the performance of the 3rd year students compared to the performance of the 4th year students /school of Arts in terms of the measure NCCT-P/T



The following table (7.6) is a summary of the figure 7.11. It groups the students who have scored within the same ratio in one column.

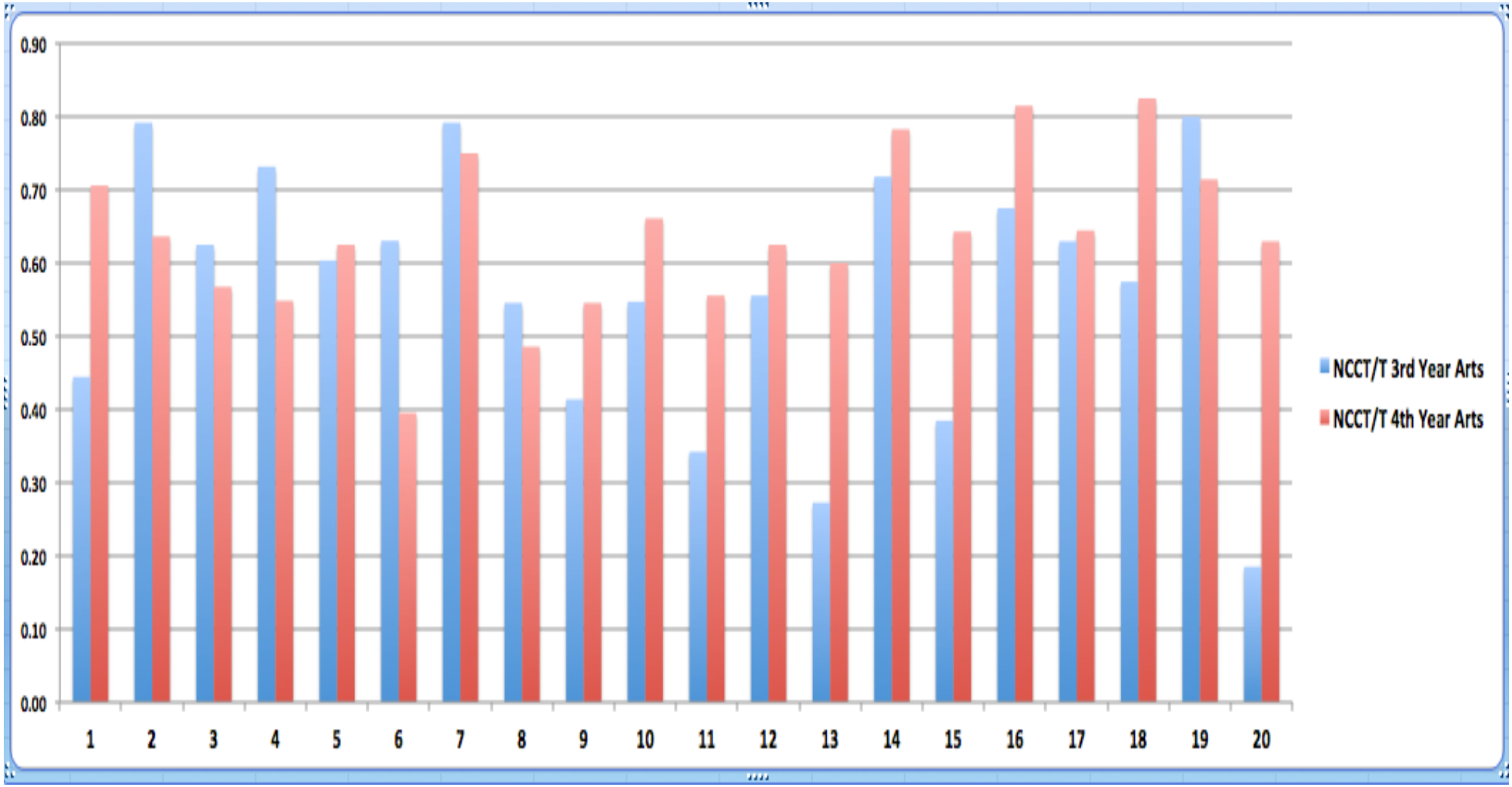
Table 7.6 the ratio of NCCT-P/T of year 3 compared to year 4/school of Arts
NCCT-P/T: A summary of figure 7.11

	0.00- 0.10	0.11-0.20	0.21-0.30	0.31-0.40	0.41-0.50	0.51-0.60	0.61-0.70	0.71-0.80	0.81-0.90	0.91-1
3rd year Arts	5	2	5	4	2	0	2	0	0	0
4th year Arts	1	2	7	3	5	0	0	2	0	0

This table (7.6) shows the ratio of the T-units that are incorrect due to sentence boundary punctuation confusion or punctuation missing errors (NCCT-P/T). Accordingly, the high ratio of NCCT-P/T depresses the performance of the students. The figure 7.11 and table 7.6 that summarizes it demonstrate that both third year and fourth year students have the problem of either placing the wrong or no punctuation at the end of the sentences. Yet it is clear that this problem is more common among the fourth year students (though the difference was not significant P=0.19). This is because only one fourth year student scored between 0.00–0.10 but five third year students made fewer errors than 0.1, and also the ratio of five of the fourth year students is between 0.41–0.50 while the ratio of only two third year students is between 0.41–0.50. Adding to this, two of the fourth year students have scored between 0.71-0.80 but only two third year students scored between 0.61-0.70.

The following figure (7.12) shows a comparison between the performance of the third year students/school of Arts and the performance of the fourth year students/school of Arts as far as the ratio NCCT /T is concerned. The horizontal axis represents the students of both years and the vertical axis represents the ratios the students have scored.

Figure 7.12 the performance of the 3rd year students compared to the performance of the 4th year students /school of Arts in terms of the measure NCCT/T



The following table (7.7) is a summary of the figure 7.12. It groups the students who have scored within the same ratio in one column.

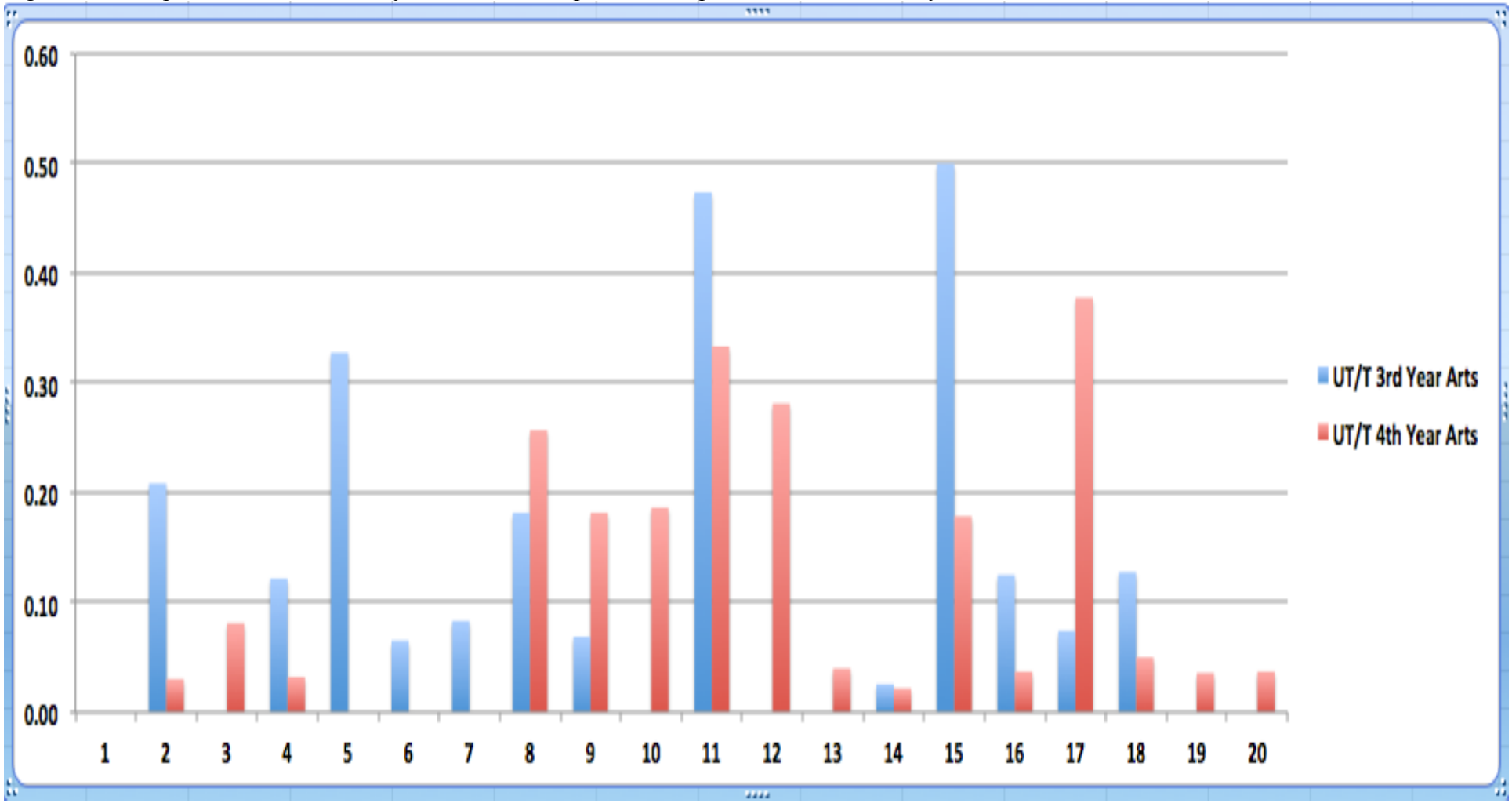
Table 7.7 the ratio of NCCT/T of year 3 compared to year 4/school of Arts
NCCT/T: A summary of figure 7.12

	0.00- 0.10	0.11-0.20	0.21-0.30	0.31-0.40	0.41-0.50	0.51-0.60	0.61-0.70	0.71-0.80	0.81-0.90	0.91-1
3rd year Arts	0	1	1	2	2	5	4	5	0	0
4th year Arts	0	0	0	1	1	5	7	4	2	0

This ratio is a measure of the T-units that are incorrect but still analyzable in terms of error analysis or one can say only 'partially' correct. It seems from the figure 7.12 and the tabulated summary in table 7.7 that the students in both groups (third year and fourth year) have produced a high number of these NCCTs. However, even within this category, the third year students appear to have outperformed the fourth year students as more fourth year students have scored higher than the third year students. No fourth year student, for example, has scored less than 0.10, 0.20, and 0.30, and only one has scored between 0.31 and 0.40 but four third year students scored within these ranges. In addition to this, seven fourth year students have obtained scores above 0.6, four obtained a ratio higher than 0.7 and two have even scored higher than 0.8 while only four third year students have reached the level above 0.6, and no students has scored above 0.8. The difference between the two seems to be close to significant as the P value of the T test =0.07.

This figure (7.13) shows a comparison between the performance of the third year students/school of Arts and the performance of the fourth year students/school of Arts as far as the ratio UT/T is concerned. The horizontal axis represents the students of both years and the vertical axis represents the ratios the students have scored.

Figure 7.13 the performance of the 3rd year students compared to the performance of the 4th year students /school of Arts in terms of the measure UT/T



The following table (7.8) is a summary of the figure 7.13. It groups the students who have scored within the same ratio in one column.

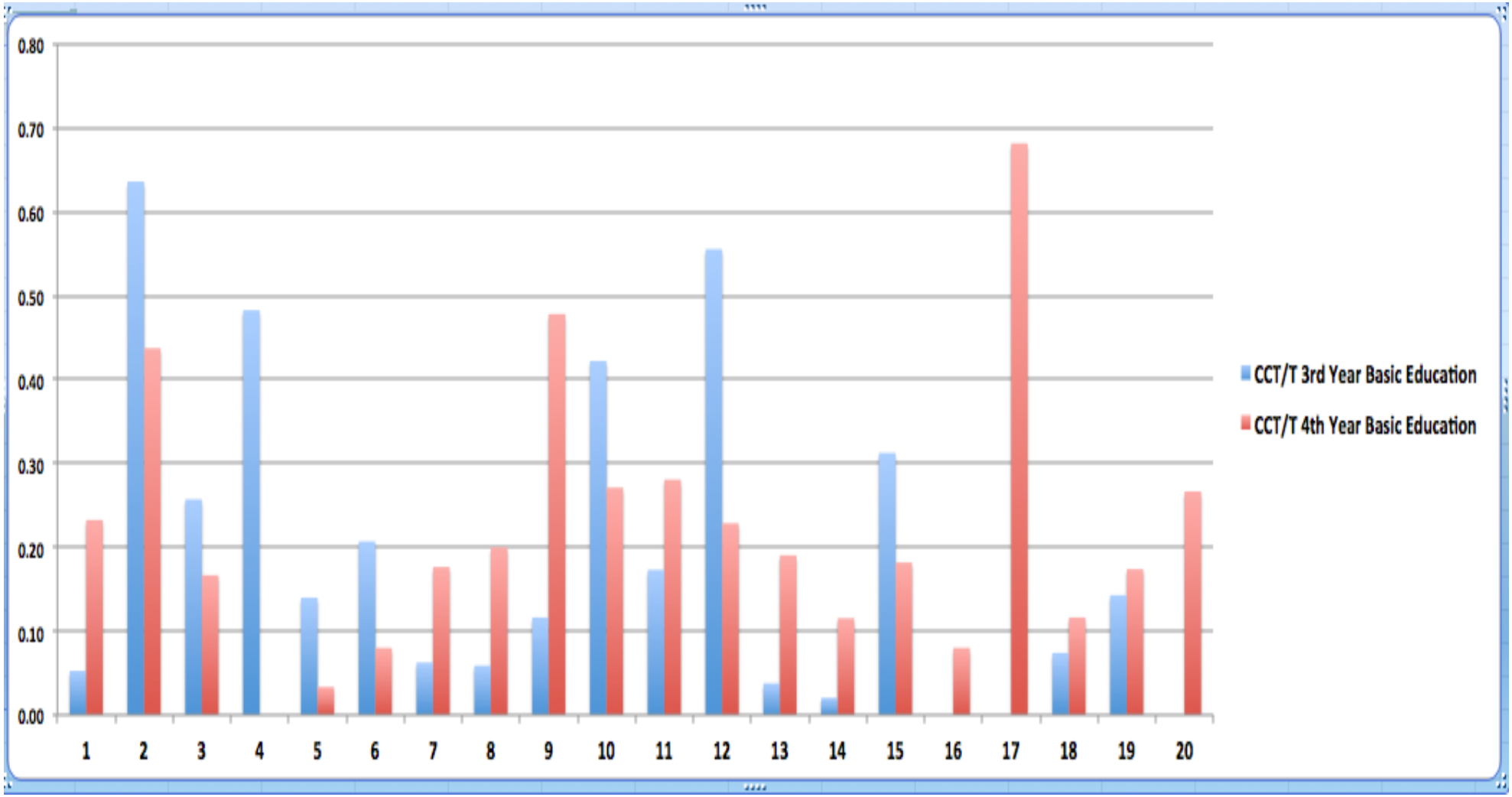
Table 7.8 the ratio of UT/T of year 3 compared to year 4/school of Arts
UT/T: A summary of figure 7.13

	0.00- 0.10	0.11-0.20	0.21-0.30	0.31-0.40	0.41-0.50	0.51-0.60	0.61-0.70	0.71-0.80	0.81-0.90	0.91-1
3rd year Arts	12	4	1	1	2	0	0	0	0	0
4th year Arts	13	3	2	2	0	0	0	0	0	0

The UT/T is a ratio of the T-units that are unanalyzable in terms of error analysis. The figure (7.13) and its summary in table 7.8 show that both third and fourth year students have scored low on this measure and there is not a very big difference between their scores. It is important to state, though, that more of the third year students have scored higher on this category, meaning that they have produced more of these UTs. As is obvious from the table 7.8, thirteen fourth year students have scored less than 0.1 but twelve third year students have scored below this ratio. In addition to this, despite the fact that the further you move to the right side of the table the number of the third and fourth year students decreases, two third year students have scored above 0.40. The difference was tested statistically and it became apparent that it was not significant (P=0.76).

This figure (7.14) shows a comparison between the performance of the third year students/ school of Basic Education and the performance of the fourth year students/school of Basic Education as far as the ratio CCT/T is concerned. The horizontal axis represents the students of both years and the vertical axis represents the ratios the students have scored.

Figure 7.14 the performance of the 3rd year students compared to the performance of the 4th year students /school of Basic Education in terms of the measure CCT/T



The following table 7.9 is a summary of the figure 7.14. It groups the students who have scored within the same ratio in one column.

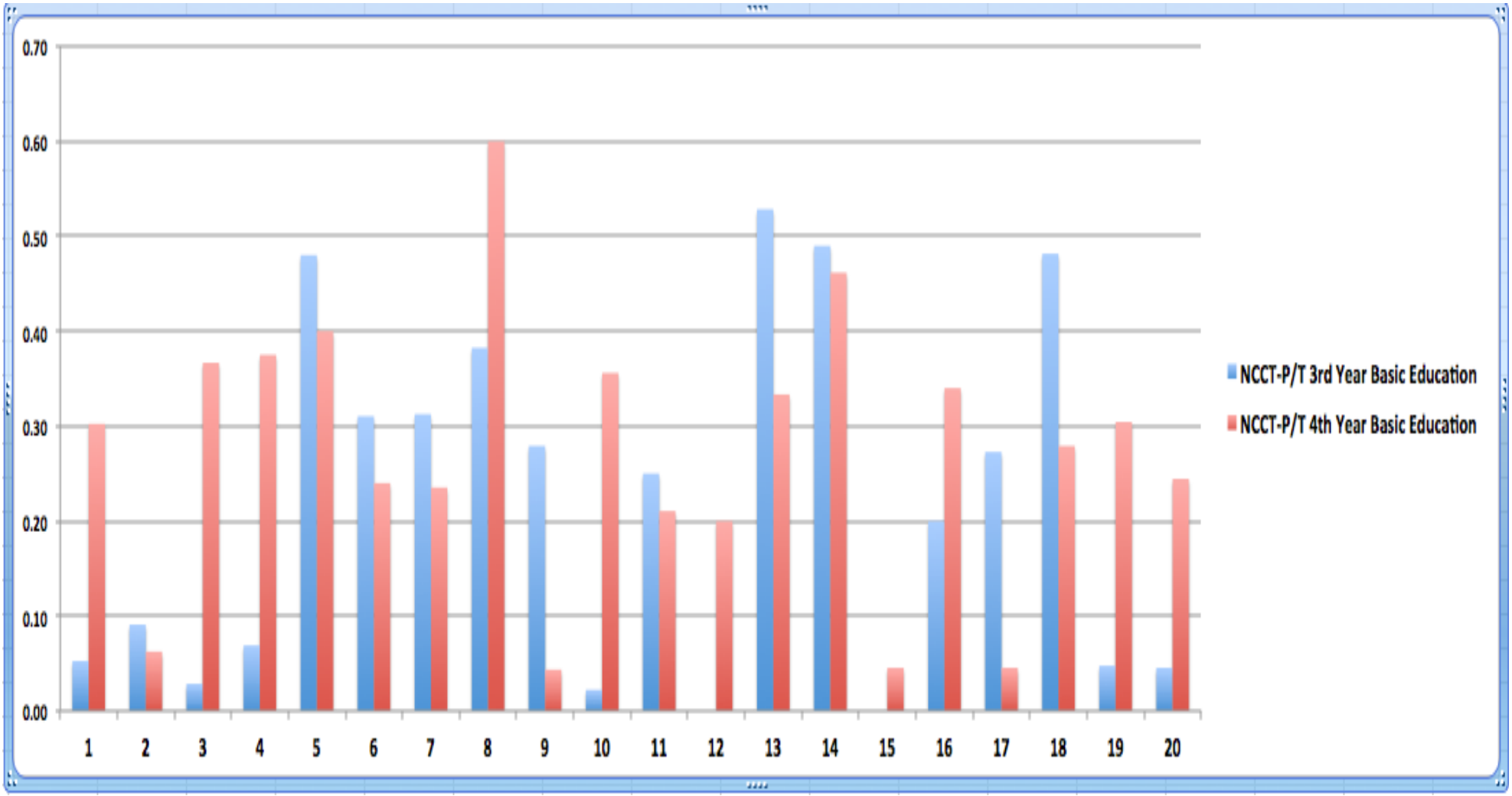
Table 7.9 the ratio of CCT/T of year 3 compared to year 4/school of Basic Education
CCT/T: A summary of figure 7.14

	0.00- 0.10	0.11-0.20	0.21-0.30	0.31-0.40	0.41-0.50	0.51-0.60	0.61-0.70	0.71-0.80	0.81-0.90	0.91-1
3rd year Basic Education	9	4	2	1	2	1	1	0	0	0
4th year Basic Education	4	8	5	0	2	0	1	0	0	0

The figure 7.14 and table 7.9 show the ratio of CCT/T. It is obvious that both the third year and fourth year students are not so different. The number of students who have scored between 0.31 and 0.9 are actually very similar and in many cases the same. The remarkable difference only lies in the score range 0.00– 0.10 to 0.21–0.30. More third year students scored below 0.10 but the number of fourth year students who have scored below 0.20 and 0.30 is higher (13 students compared to 6 for third year students). This could indicate that the fourth year students have slightly outperformed the third year students especially on the lower ratio levels. This is also evident from the fact that the number of third year students decreases as one moves to the right side of the table but the number of fourth year students increases. A T -test of the result has resulted in a non-significant difference (P=0.57).

This figure (7.15) shows a comparison between the performance of the third year students/ school of Basic Education and the performance of the fourth year students/school of Basic Education as far as the ratio NCCT-P/T is concerned. The horizontal axis represents the students of both years and the vertical axis represents the ratios the students have scored.

Figure 7.15 the performance of the 3rd year students compared to the performance of the 4th year students /school of Basic Education in terms of the measure NCCT-P/T



The following table (7.10) is a summary of the figure 7.15. It groups the students who have scored within the same ratio in one column.

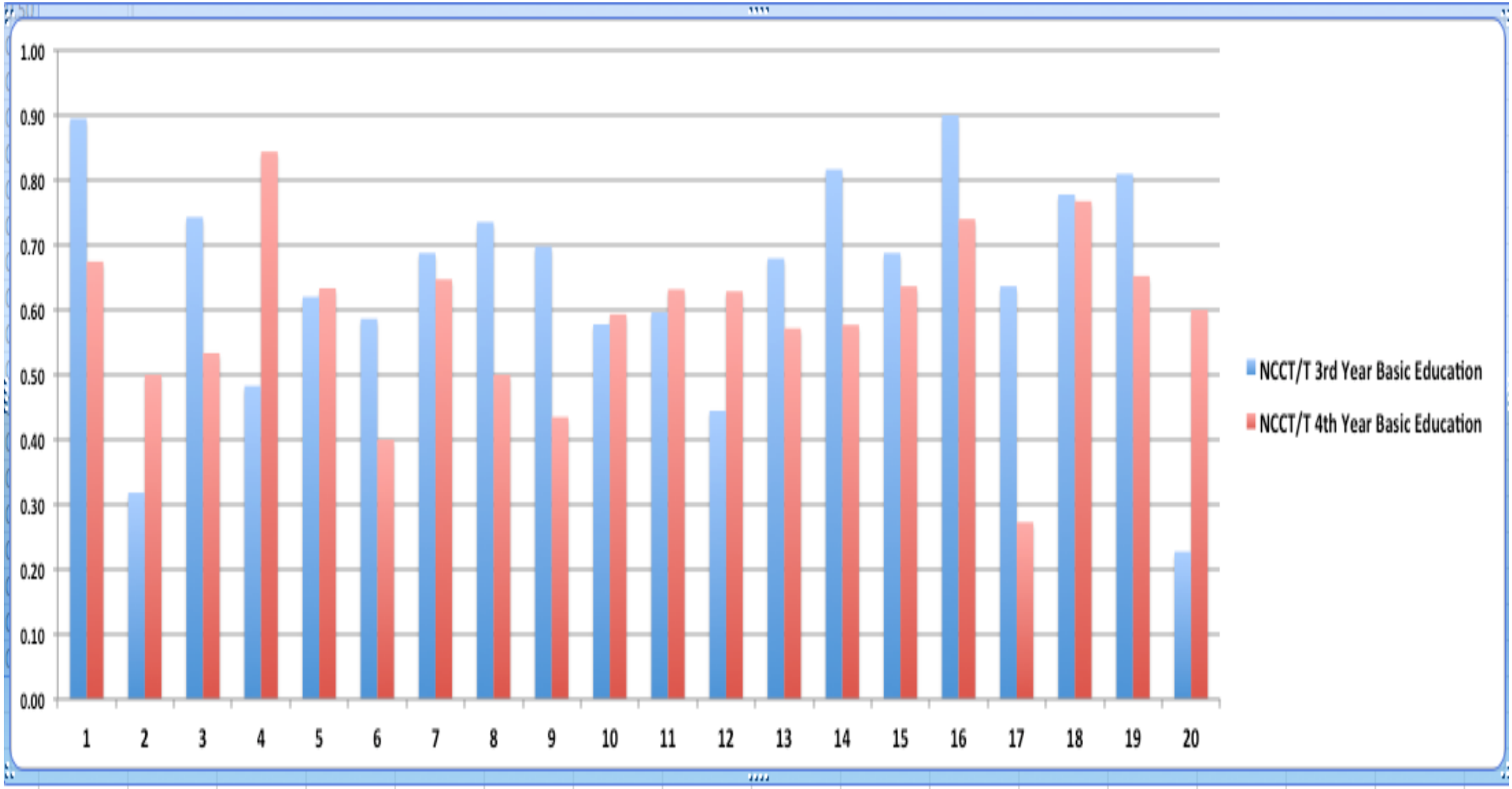
Table 7.10 the ratio of NCCT-P/T of year 3 compared to year 4/school of Basic Education
NCCT-P/T: A summary of figure 7.15

	0.00- 0.10	0.11-0.20	0.21-0.30	0.31-0.40	0.41-0.50	0.51-0.60	0.61-0.70	0.71-0.80	0.81-0.90	0.91-1
3rd year Basic Education	9	1	3	3	3	1	0	0	0	0
4th year Basic Education	4	1	7	6	1	1	0	0	0	0

Looking at table 7.10 and figure 7.15, one can conclude that both third and fourth year students have committed errors in punctuation at sentence boundary. As is the case with the school of Arts, more fourth year students have scored high on this measure. This is clear from the table (7.10) because the number of third year students is high in the low score categories (9 students scored below 0.1) and the further one moves towards the higher scores, the lower the number of students becomes. On the contrary, the number of fourth year students increases the further one moves to the right side of the table.

This figure (7.16) shows a comparison between the performance of the third year students/ school of Basic Education and the performance of the fourth year students/school of Basic Education as far as the ratio NCCT/T is concerned. The horizontal axis represents the students of both years and the vertical axis represents the ratios the students have scored.

Figure 7.16 the performance of the 3rd year students compared to the performance of the 4th year students /school of Basic Education in terms of the measure NCCT/T



The following table (7.11) is a summary of the figure 7.16. It groups the students who have scored within the same ratio in one column.

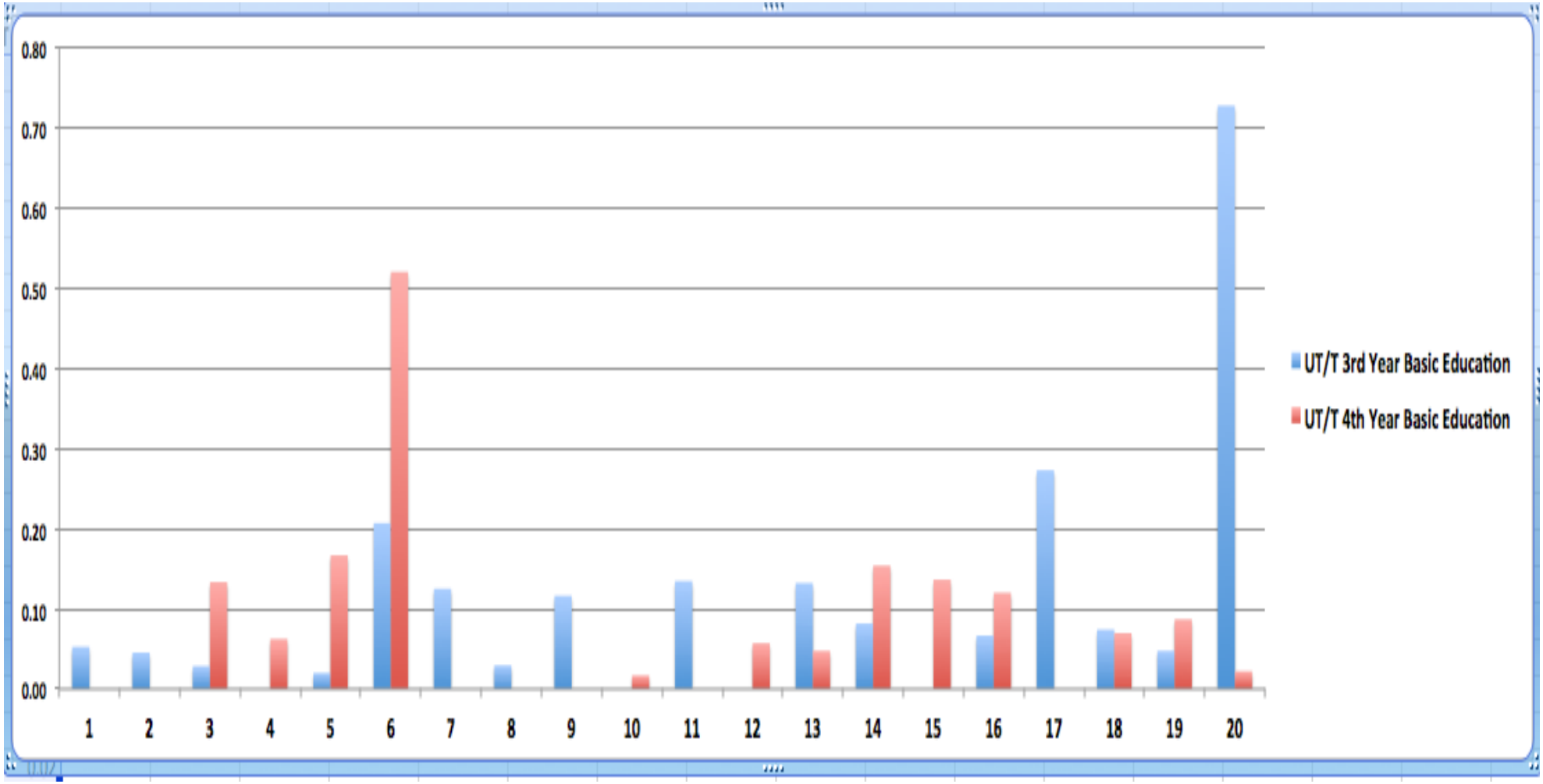
Table 7.11 the ratio of NCCT/T of year 3 compared to year 4/school of Basic Education
NCCT/T: A summary of figure 7.16

	0.00- 0.10	0.11-0.20	0.21-0.30	0.31-0.40	0.41-0.50	0.51-0.60	0.61-0.70	0.71-0.80	0.81-0.90	0.91-1
3rd year Basic Education	0	0	1	1	2	3	6	3	3	1
4th year Basic Education	0	0	1	1	3	5	7	2	1	0

As with the case of the students of the school of Arts, both third and fourth year students of the school of Basic Education (figure 7.16 and table 7.11) have scored high on the ratio of NCCT/T. This indicates that their performance contains a lot of these T-units that are only incorrect because of some minor errors or few major errors. The numbers of the fourth and third year students in the score range between 0.00–0.10 to the range 0.31–0.40 are the same. The difference nonetheless appears to start at the score range 0.41-0.50, where fourth year students are more in number than third year students (3 compared to 2). Also, more fourth year students have scored higher than 0.6 (7 compared to 6 third year students). A higher number of third year students, however, have scored within the range 0.71 to 1 within the ranges (7 compared to 3 fourth year students). The difference, however, is not statistically significant (P=0.24).

This figure (7.17) shows a comparison between the performance of the third year students/ school of Basic Education and the performance of the fourth year students/school of Basic Education as far as the ratio UT/T is concerned. The horizontal axis represents the students of both years and the vertical axis represents the ratios the students have scored.

Figure 7.17 the performance of the 3rd year students compared to the performance of the 4th year students /school of Basic Education in terms of the measure UT/T



The following table (7.12) is a summary of the figure 7.17. It groups the students who have scored within the same ratio in one column.

Table 7.12 the ratio of UT/T of year 3 compared to year 4/school of Basic Education
UT/T: A summary of figure 7.17

	0.00- 0.10	0.11-0.20	0.21-0.30	0.31-0.40	0.41-0.50	0.51-0.60	0.61-0.70	0.71-0.80	0.81-0.90	0.91-1
3rd year Basic Education	13	4	2	0	0	0	0	1	0	0
4th year Basic Education	14	5	0	0	0	1	0	0	0	0

The ratio of the unanalyzable T-units that are produced by the third and fourth year students at the school of Basic Education is, as is clear from the figure 7.17 and table 7.12, not so high. This is similar to the ratios scored by the third and fourth year students at the school of Arts (figure 7.13 and table 7.8). The performance of more fourth year students seems to be better than the performance of the third year students (14 fourth year students scored below 0.1 compared to 13 of the third year students). Moreover, 4 third year students have scored below 0.2 compared to 5 of the fourth year students. One third year student has even scored above 0.70. The difference has been tested statistically and it is not significant (P=0.53).

7.6 Conclusion

This chapter represents a detailed presentation of a new method for measuring accuracy in second and foreign language writing. The method is data driven i.e. it depends to a great extent on the nature of the data dealt with. In other words, I have devised the method to be able to objectively assess the type of essays I have taken as a sample. I have called the method 'An Integrated Approach to Achievement' (IAA) because it integrates different methods and looks more at the achievement of the students rather than their failure. This method combines error analysis, T-units-based correctness analysis and various-units-based correctness analysis that have been applied in the previous chapters. It uses the T-unit as a basic unit of analysis, and first distinguishes T-units from fragments. It is also based on a 'three-category' principle where it differentiates among three types of T-units, namely CCT, NCCT and UT. Moreover, it includes another category of T-units that are incorrect because of sentence boundary punctuation. Nevertheless, this category might be used to label NCCTs and UTs or it might be used as a category by itself. For example, a T-unit might be labeled as both NCCT and NCCT-P or it might be given two labels as UT and NCCT-P. The NCCTs and UTs are further analyzed in terms of error analysis and various-units-based correctness analysis respectively. With error analysis, the types and numbers of errors are identified in the NCCTs and with the

various-units-based correctness analysis, the UTs are divided into the small correct and almost correct stretches. In this way, this method gives us an idea about the internal structures of both NCCTs and UTs; for example, how many and how severe the errors in the NCCTs are and to what extent the UTs are structured.

The method has been applied to a sample of the data collected for the present study leading to a number of important findings. One of these findings is that the method can strictly place students into different levels as far as the accuracy of their essays is concerned. The second finding is that, contrary to my expectation that the majority of T-units would be categorized as UTs, the students have produced a small number of these UTs. This shows us that reading an essay and paying attention to its minute details might be much better for assessment than merely giving it a grade by only considering its assessment holistically. The third finding is that the third year students and fourth year students in both schools (school of Arts and school of Basic Education) tend to be similar in their production in terms of all ratios (CCT/T, NCCT/T NCCT-P/T, UT/T) with some minor differences (though not statistically significant) between the third year students and fourth year students in both schools. This actually points to something important; development in writing skills does not seem to depend very much on the year of study.

The method differs from error analysis in that it considers correct forms as well as errors. It is different from the T-unit and clause-based correctness analysis as it accounts for error gravity in T-units, and it is different from the various-units-based correctness analysis as it considers one type of unit and breaking down the UTs into small stretches that are either correct or almost correct.

Notes

1. The other 4 are also NCCTs besides being NCCT-P.

Chapter Eight

Measurement of Syntactic Complexity in Second Language Writing

8.1 Introduction

Complexity is one dimension in the triangle of measuring language proficiency, alongside accuracy and fluency. Syntactic complexity is one type of linguistic complexity (see figure 8.1 below) and this chapter is a description of this construct. The chapter first defines complexity and explains a number of its measures that have been used in previous research. The chapter then demonstrates the application of a number of these measures and a number of new measures to a sample of the data from the present study.

This chapter aims to answer the following questions: (1) What are the measures used in measuring syntactic complexity in L2 writing? (2) Can a number of them be applied to the data of the present study? What other measures can also be used? (3) To what extent do the *accuracy* scores obtained in the previous chapter differ from or resemble the *complexity* scores of the same students, (within-class comparison)? (4) How do the argumentative and narrative essays written by the students of the same year differ from or resemble each other in terms of complexity (within-class comparison)? (5) How does the complexity of the 3rd year students' narrative essays differ from or resemble the complexity of the 4th year narrative essays (between-class comparison)? (6) How does the complexity of the 3rd year students' argumentative essays differ from or resemble the complexity of the 4th year argumentative essays (between-class comparison)? (7) Is increased complexity a sign of improvement? (8) Is phrasal complexity more difficult for students than clausal complexity, or vice versa?

Bearing in mind Lambert and Kormos' (2014) argument that it is better not to measure subordination as one construct but to measure its structures separately, I have taken different measures of complexity such as subordination (e.g. 8.1 S(2) *Also I tried to make my relationship with my teachers as good as I could <u>because they would be my right hand for encourage me and develope my language from the worst to the best</u>*) and coordination (e.g. 8.2 S(8) *Finally, I came to conclude my Essay<u>, and I going to say</u> that I was so intreasted in week in my college as a student<u>, and I prode of that, and I am going to continue my Education in college</u>*.) Also, I have measured the ratio of non-finite clauses and phrases as nominal and adverbials (e.g. 8.3 S(33) *I think <u>living in the country side</u> is more healthy for people*…, e.g. 8.4 S(139) *<u>Thinking of that the next hour I have a rest</u>, I spent the whole hour in her office*)

separately from subordination. Moreover, Norris and Ortega (2009) have called for the use of different measures that operationalize subordination, coordination and phrasal complexity. More importantly, Biber *et al.* (2011) have pointed out that an increase in phrasal complexity rather than clausal complexity is one of the characteristics of academic writing. It is also important that I have found it very necessary to measure phrasal complexity through measuring postmodification (including different postmodifiers such as prepositional phrases e.g. 8.5 S(254) *in the process of learning English language*, relative clauses e.g. 8.6 S(258) *Is a journey that human can not forget it*, *to-infinitive* and participial phrases e.g. 8.7 S(37) *the main reason to pollute the environment.* e.g. 8.8 S(270) *cigeratte is has a material called (Necotin) that is affected in lungs*) because I believe, based on my experience as a teacher of EFL, an increase in postmodification especially the use of participial clauses and phrases is a type of complexity that is most indicative of sophistication in language. Taking all this in to consideration, I have used a number of new measures of phrasal and clausal complexity that are described in section 8.3.2.1

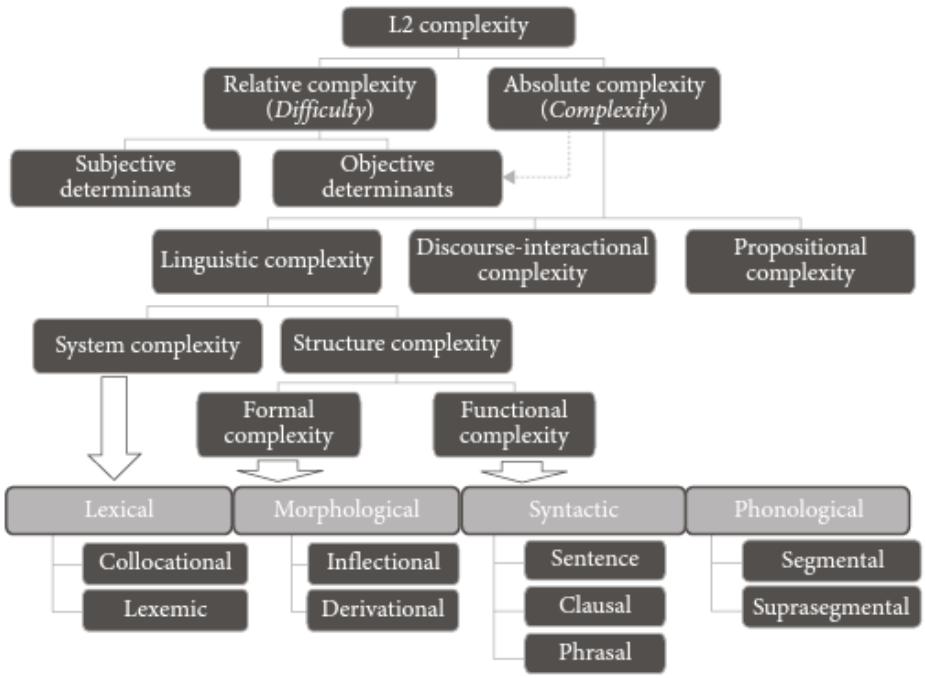
Section 8.2 in this chapter gives a definition of complexity in general, describing its types, and listing and explaining the ways of measuring syntactic complexity. Section 8.3 is allocated to how the measures of syntactic complexity that are used in this chapter work. This section describes the sample of the data considered for this chapter, the measures used in data analysis and the way these measures are applied to the data (i.e. dividing the text into units). Section 8.4 includes a description of the process of engaging another user in the analysis of a section of the data. Section 8.5 describes the detailed application of the method of analysis and the measures of syntactic complexity to one essay. Section 8.6 presents the results and discussion of the application of the measures of complexity that are used in this chapter to a sample of the data of the present study. In this section, various comparisons are carried out. First (subsection 8.6.1), the accuracy scores of ten essays analyzed in the previous chapters are compared to their complexity scores. Then in (subsection 8.6.2), a comparison is carried out between the complexity scores of the narrative essays to the complexity scores of the argumentative essays written by the students of the same year. In other words, these two subsections (8.6.1 and 8.6.2) are allocated to a within-class comparison. The subsections 8.6.3 and 8.6.4 are allocated to between-class comparison comparing essays of the same type of different years of study to each other i.e. for example, the narrative essays of the third year students to the narrative essays of the fourth year students of the same school. Section 8.7 is a conclusion of what has been discussed in this chapter.

8.2 Complexity: definition and measurement

8.2.1 Definition

Complexity has been characterized as a difficult construct to define and operationalize (Housen and Kuiken, 2009) and hence "ill-defined" in the literature (Bulté and Housen, 2014:45). One of the reasons why complexity is "ill-defined", as argued by Bulté and Housen (ibid), is that it has not been investigated as a "central" variable. Many studies use complexity as a variable to measure the effect of other factors like age, various types of instruction, specific second/foreign language programme, task planning etc. on L2 language development and L2 proficiency and to describe the complexity features of genres like narratives or modes of language (written vs. spoken language) (see, for example, Lintunen and Mäkilä, 2014; Seyyedi *et al.* 2013; Johnson *et al.* 2012; Vyatkina, 2012; Kormos, 2011; R. Ellis and Yuan, 2004; Yuan and R. Ellis, 2003). The second reason why complexity has not been precisely defined in the literature is its intricate nature where different layers and types are interconnected in one single framework making it difficult for researchers to define what they are actually measuring. Does "more complex" mean "difficult to acquire or to produce," "acquired late(r)," "developmentally advanced," "more proficient," "more mature," "of high(er) quality," or simply "better" ? (Bulté and Housen 2014:45). Bulté and Housen (2012:23, see also Bulté and Housen, 2014:44) have provided a complex picture of complexity as a construct:

Figure 8.1 the taxonomy of complexity put forward by Bulté and Housen, (2012:23)



According to this figure (8.1), complexity has different types. The first two main types are relative complexity and absolute complexity or as Bultè and Housen, (2014:43) name them "cognitive complexity" (difficulty) and "absolute complexity" (or simply "complexity"). Cognitive complexity is "the relative difficulty with which language features are processed in L2 performance and acquisition" (Housen and Kuiken, 2009:463). In other words, cognitive complexity is looking at any structure or item from the L2 learner's perspective (ibid). This complexity is thus "relative" and can be different from one learner to another. For example, a learner may find it easy to learn tenses and another learner may find it very difficult. Or, one structure can be learned earlier than others and hence psycholinguistic research is needed to identify which linguistic structures are cognitively difficult to learn. However, the ease or difficulty with which a learner learns a given structure might or might not be due to the inherent complexity of the structure. The absolute or inherent complexity of the structure is "the number of discrete components that a language feature or a language system consists of, and as the number of connections between the different components." (Bultè and Housen, 2012:24)

Other scholars have not made this distinction very clear in their definition. An example of this are the definitions of complexity given by R. Ellis and Barkhuizen (2005) as either referring to the learner's "willingness to use more challenging and difficult language" or the "learner's preparedness to use a wide range of different structures." (R. Ellis and Barkhuizen, 2005:139)

Absolute complexity is further divided into propositional complexity, discourse-interactional complexity and linguistic complexity. While propositional complexity (how many ideas are expressed in a task) and discourse-interactional complexity (the way two or more interlocutors interact, for example, in terms of turns) have not received much attention, linguistic complexity has been the center of a considerable amount of research (Bultè and Housen, 2014).

The study of linguistic complexity can be either the study of the learner's "global and dynamic" system of language (system complexity) or the study of the stable properties of the discrete components (items and structures) that make up this system (structure complexity) (Bultè and Housen 2012:25). Bultè and Housen (ibid) define system complexity as "the degree of elaboration, the size, breadth, width or richness of the learner's second language system or repertoire" and structure complexity as the complexity that "has to do more with the depth than the breadth or range." To be more specific a researcher might study the form or the functional complexity of an item or a structure and this can be on different levels i.e. lexical, morphological, syntactic and phonological.

Taking the above discussion into consideration, the notion of syntactic complexity seems to be very complex to define. For example, it might be thought that the number of clauses in a sentence defines syntactic complexity, or the extent to which learners use subordination. Other definitions might incorporate notions of transformations (for example passive might be more complex than active or reduced relative clauses might be more complex than the relative clauses with clause markers). However, it is beyond the scope of this research to arrive at a definition for syntactic complexity; what I am interested in instead is finding ways of measuring it and the following is a list of the structures that operationalizes syntactic complexity in this research:

- long T-units,
- subordination,
- coordination,
- non-finite ing-clauses and phrases in nominal and adverbial position,
- *to-infinitive* phrases and clauses in nominal position, and
- noun phrases with postmodification in terms of prepositional phrases, *to-infinitive* phrases, relative clauses and participial clauses.

I have particularly chosen non-finite clauses and phrases as adverbial and nominal because first, they themselves are marked structures (as opposite to the unmarked finite structures) and second they are embedded. Moreover, their presence or absence can convey information about clause complexity. The choice of postmodification is due to the fact that phrasal complexity, especially as far as postmodification is concerned, is not researched very widely and, as mentioned earlier, it is the sign of a high sophistication in language.

8.2.2 Measurement of complexity

Different measures have been used to measure the syntactic complexity of L2 language writing and speaking. Bultè and Housen (2014), for example, have listed the following measures (in table 8.1) which include measures that have been used in forty research studies and some others suggested by Bultè and Housen (2014) themselves (see also Iwashita, 2006, for measures of oral proficiency of English learners learning Japanese and R. Ellis, 2003 for measures of fluency, accuracy and complexity):

Table 8.1 the measures of syntactic complexity that are listed in Bultè and Housen[1] (2012:30-31)

	Syntactic						
	Overall	**Sentential-Coordination**	**Sentential-Subordination**	**Sentential (Clausal and Phrasal)**	**Clausal**	**Phrasal**	**Other (Syntactic sophistication)**
1	Mean length of T unit	Coordinated clauses/clauses	Clauses/AS-unit	Mean length of clause	Syntactic arguments/clause	Dependents/(noun, verb) phrase	Frequency of passive forms
2	Mean length of C-unit		Clauses/C-unit	S-nodes/clause			Frequency of infinitival phrases
3	Mean length of turn		Clauses/T-unit				Frequency of conjoined clauses
4	Mean length of AS-unit		Dependent clauses/clauses				Frequency of Wh-clauses
5	Mean length of utterance		Number of subordinate clauses				Frequency of imperatives
6	S-nodes/ T-unit		Subordinate clauses/clauses				Frequency of auxiliaries
7	S-nodes/AS-unit		Subordinate clauses/dependent clauses				Frequency of comparatives
8			Subordinate clauses/T – units				Frequency of conditionals
9			Relative clauses/T- units				
10			Verb phrases/T – unit				

As with accuracy measures, Wolfe-Quintero *et al.* (1998:99) have also provided a usefully descriptive list of complexity measures that have been used in different studies, which yields different results. Among the measures that they described as valid to be used as metrics for language development are Clause/T-units, Dependent clauses/Clauses and Dependent clauses/T-units because "they exhibit a linear relationship to proficiency level across studies that used a wide range of levels."

Some researchers have used a number of measures for an automated analysis of the data. For example, Lu and Ai (2015:18, see also Ai and Lu, 2013:254) have used 14 measures to measure the syntactic complexity of a number of essays written by non-native speakers with different L1 backgrounds extracted from the International Corpus of Learner English (ICLE Granger, Dagneaux, Meunier, and Paquot, 2009) compared to the essays written by native speakers (in the Louvain Corpus of Native English Essays (LOCNESS) Granger, 1996) using the L2 Syntactic Complexity Analyzer[2] that is designed with these 14 measures as a package. The following (table 8.2) are the measures:

Table 8.2 the syntactic measures in the Syntactic Complexity Analyzer

Measure	Code	Definition
Length of production unit		
Mean length of clause	MLC	# of words/# of clauses
Mean length of sentence	MLS	# of words/# of sentences
Mean length of T-unit	MLT	# of words/# of T-units
Amount of subordination		
Clauses per T-unit	C/T	# of clauses/# of T-unit
Complex T-units per T-unit	CT/T	# of complex T-units/# of T-units
Dependent clauses per clause	DC/C	# of dependent clauses/# of clauses
Dependent clauses per T-unit	DC/T	# of dependent clauses/# of T-units
Amount of coordination		
Coordinate phrases per clause	CP/C	# of coordinate phrases/# of clauses
Coordinate phrases per T-unit	CP/T	# of coordinate phrases/# of T-units
T-units per sentence	T/S	# of T-units/# of sentences
Degree of phrasal sophistication		
Complex nominals per clause	CN/C	# of complex nominals/# of clauses
Complex nominals per T-unit	CN/T	# of complex nominals/# of T-units
Verb phrases per T-unit	VP/T	# of verb phrases/# of T-units
Overall sentence complexity		
Clauses per sentence	C/S	# of clauses/# of sentences

An automated analysis of the corpus of the present study would have been a very good experience if that had been possible. The L2 Syntactic Complexity Analyzer was developed for the corpus produced by advanced college level students (Lu, 2010, see also Lu, 2011). Most of the corpus of the present study, though by university students, is of a low level that could not be subjected to an automated analysis.

8.3 The application of complexity measures to a sample of the data of the present study

A number of sentential, clausal and phrasal complexity measures have been applied to 120 essays as a sample selected from the data of the present study. The following subsections comprise a description and explanation of the sample chosen, the type of measures used, the way the data were divided into units, the description of the process of engaging a second rater in the analysis of the data, the detailed application of the method to one essay as a sample, and the results and discussion of the application of the method and measures to the selected sample.

8.3.1 the sample

The sub-corpus subjected to this analysis consists of 120 essays. These essays have been written by the students of three schools of two universities (School of Arts, and School of Basic Education/ University of Dohuk) whose data have previously been used for the analysis of accuracy, and School of Languages/ University of Zakho whose data are used for the first time in this analysis of complexity). 20 essays per each year per each school (20×6=120) have been selected. Each block of 20 essays comprises 10 narrative and 10 argumentative essays. That means the sub-corpus is divided into 60 narrative and 60 argumentative essays. Forty of these (narrative) essays of the students in the two schools of Arts and Basic Education have been analyzed for accuracy in the previous chapters. These 40 essays consist of 10 essays per each year per each school (of Arts and Basic Education). This can be better shown in the following table:

Table 8.3 the sample used in this chapter

School of Arts/ University of Dohuk				School of Basic Education/University of Dohuk				School of Languages/ University of Zakho			
3rd year		4th year		3rd year		4th year		3rd year		4th year	
N	A	N	A	N	A	N	A	N	A	N	A
10 PU	10	10 PU	10	10 PU	10	10 PU	10	10	10	10	10

N = narrative A= argumentative PU= previously used in the chapters on accuracy.

8.3.2 Description of measures and dividing the text into units

8.3.2.1 The measures

A number of measures have been used for measuring the complexity of the sampled essays. Some existing measures have been used but some others are used for the first time, such as the postmodification measures and the non-finite clauses and phrases per/T-unit. The measures are as follows (table 8.4):

Table 8.4 the measures of syntactic complexity used in this chapter

Areas to be measured	*Overall complexity*	*Coordination*	*Subordination and embeddedness*	*Phrasal complexity*	*Clausal complexity (Measures of to- infinitive phrases functioning as nominal and ing and ed-clauses and phrases functioning as nominal and adverbials)*
Measures	Mean length of T unit (W/T words per T units)	Total number of T units per total number of sentences (T/S)	Dependent and embedded clauses/T unit (DC/T)	Total number of prepositional phrases as post nominal modifiers/Total number of noun phrases (PP/NP)	Non-finite clauses/phrases as nominal and adverbials / total number of T units (NF Clauses, Phrases/T)
				Total number of relative clauses / Total number of noun phrases (RC/NP)	
				Total number of *to- infinitive phrases* as post nominal modifiers/ total number of noun phrases (To-inf./ NP)	
				Total number of participial clauses and phrases (*ing-participles and ed-participles*) as post nominal modifiers/ Total number of noun phrases (P/NP)	

8.3.2.2. Identifying relevant units

1. Dividing the text into sentences and T-units

In order to apply these measures, the essays are first cut up into T-units and sentences as basic units of analysis for measuring sentential and overall complexity. The T-unit is defined (as in chapter five) as "one main clause plus whatever subordinate clauses … attached to that main clause" (Hunt, 1966:737) but the sentence can be defined here as a string of words that are *syntactically eligible* to start with a capital letter at the beginning of its first word and end with a sentence ending mark like a period, a question or an exclamation mark. It is very important here to note that the sentence might consist of a number of T-units (no matter how many) that are coordinated by any coordinating conjunctions like *and*, *but*, *or* etc. Notice the following examples produced by different subjects:

e.g 8.9 S(167)

[The first of all, the cigaratte is something not good for health(T1) *and it causes ills,* (T2) *and everybody should be prevent from it* (T3)] one sentence.

e.g. 8.10 S(250)

[Fainaly I thing I am not the only person who had across in this experaince (T1)*, but I am one of them*(T2) *and I want to write my story and my journey from Mosul to Zakho and what was the reson behind that journey* (T3) *and I always thank god because he save my and my family from anything bad happend to us*(T4)] one sentence

It is clear that the text in example 8.9 consists of three T-units but only one sentence, and in the example 8.10, the text comprises four T-units and only one sentence.

However, if there is a full stop or any other sentence end mark preceding the coordinating conjunction, the string of words starting with this coordinating conjunction will be considered a separate sentence. Yet, with a comma before the coordinating conjunction, the string of words following this coordinating conjunction will remain as a part of the preceding sentence. The following are examples of this point:

e.g. 8.11 S(122)

I have found out that our house was the safest place ever. and it was only two weeks away when I had been beaten up by two boys and I almost had my pinkie broken and I retaliated by hitting him by stone on the head. and subsequently I got kicked out from school for two weeks before been forced to sign up for not breaking the rules again.

Table 8.5 the division of a sample paragraph into T-units and sentences

	T-units	Sentences
1	*I have found out that our house was the safest place ever. (T1)*	*I have found out that our house was the safest place ever. (S1)*
2	*and it was only two weeks away when I had been beaten up by two boys (T2)*	*and it was only two weeks away when I had been beaten up by two boys and I almost had my pinkie broken and I retaliated by hitting him by stone on the head. (S2)*
3	*and I almost had my pinkie broken (T3)*	*and subsequently I got kicked out from school for two weeks before been forced to sign up for not breaking the rules again. (S3)*
4	*and I retaliated by hitting him by stone on the head. (T4).*	
5	*and subsequently I got kicked out from school for two weeks before been forced to sign up for not breaking the rules again (T5).*	

The text in example 8.11 consists of 5 T-units and 3 sentences. Although the two strings of words *I have found out that our house was the safest place ever* and *it was only two weeks away when I had been beaten up by two boys* are coordinated by the conjunction (*and*), this *and* is preceded by a full stop. That is why the string after it has been considered a separate sentence. The same applies to the two strings *I retaliated by hitting him by stone on the head* and *subsequently I got kicked out from school for two weeks before been forced to sign up for not breaking the rules again.* Notice, however, that sentence 2 consists of three strings of words joined by the conjunction (*and*) and are not separated by a full stop or any sentence terminating end.

While in the following text (example 8.12) the string *But I encourage her that he will recover by thanking God* is considered a separate sentence from the string preceding it because it is preceded by a full stop which marks the end of the previous sentence.

e.g. 8.12 S(4)

even we didn't laugh and enjoy so we felt really sad for my father and my mother couldn't believe that my father will recover (sentence one). *But I encourage her that he will recover by thanking God* (sentence two)

It is also worth mentioning that if strings of words are separated by commas without any coordinating conjunction, they are divided as separate sentences because syntactically they are eligible to end with a sentence end mark rather than a comma.

e.g. 8.13 S(123)

The first day my sister brought me to my college, she was student in lawer college

Example 8.13 consists of 2 sentences *The first day my sister brought me to my college* and *she was student in lawer college* although they are separated only by a comma. If the two strings of words were joined by any coordinating conjunction, they would be considered one sentence rather than two even in the presence of comma.

2. Dividing the text into clauses and phrases

After dividing the texts into sentences and T-units, the dependent and embedded clauses are all identified for measuring the amount of subordination and embeddedness. The types of clauses identified are the following:

a. Finite dependent (subordinate) adverbial clauses:

e.g. 8.14 S(267)

it is a good location to have rest when you have some problems.

e.g. 8.15 S(10)

We had a very interesting week, because we went to places that I had never seen them before.

In these two sentences the underlined clauses are subordinate clauses functioning as adverbials

b. Finite dependent relative clauses:

e.g. 8.16 S(141)

One of the journey will be an unforgettable journey for the person who go to trip

e.g. 8.17 S(66)

Lets get out from those pathetic cases and situations and start with the best things that happened in my life.

e.g. 8.18 S(139)

the day *when my grandfather died was the saddest event in my life.*

The three underlined clauses are relative clauses modifying the noun preceding it.

c. Finite dependent nominal clauses

e.g. 8.19 S(252)

The very first time that I met my best friend (Bezheen) was when I began to study at the secondary school called Zakho.

e.g. 8.20 S(86)

when they believe that what is performed in the imaginary world in the real world.

e.g. 8.21 S(293)

but what made me to feel so nervous was that the time when I reached my class for the first time.

In example 8.19, the underlined clause functions as subject complement. While in example 8.20, the underlined clause functions as direct object and in example 8.21, the underlined

dependent *what clause* functions as subject. These positions are normally filled by nouns and that is why these clauses are nominal clauses.

Something that needs to be highlighted here is that when two or more clauses are embedded one into the other they are all identified, first as parts of each other and then as separate clauses.

e.g. 8.22 S(293)

many different things in this place which I owe to it in my entire life {because it tought me [how to deal with different kind of people (who you never speak to them before])} CLAUSE 3) CLAUSE 2] CLAUSE 1} *actualy through my experience in my college I have learned {how to depend on myself (when I am far from my dearly family)}* CLAUSE 4} CLAUSE 5).

In example 8.22, first the clause *because it tought me how to deal with different kind of people who you never speak to them before* is identified as clause one and then the two dependent and embedded clauses *how to deal with different kind of people who you never speak to them before* and *who you never speak to them before* are identified as clause 2 and 3 respectively. The same applies to clauses 4 and 5. In this case, the text in example 8.22 contains 5 dependent and embedded clauses, one embedded in the other.

d. Non-finite *to-infinitive* nominal clauses and phrases

e.g. 8.23 S(192)

the best way is to live in a big city

e.g. 8.24 S(228)

if you want to study in Another university

In examples 8.23 and 8.24, the *to-infinitive* clauses function as nominal (subject complement in example 8.23 and direct object in example 8.24)

e. Non-finite *ing*-clauses and phrases functioning as nominal

e.g. 8.25 S(260):

I started thinking about my future and out some plans for my future.

e.g. 8.26 S(201)

ozone prevent harmful Radiation from the sun from reaching the earth.

e.g. 8.27 S(142)

Knowing culture is another important thing in a journey to have happiness.

The underlined *ing-clauses* function as direct object, object of preposition, and subject in examples 8.25, 8.26, and 8.27 respectively.

f. Non-finite *ing*-clauses and phrases functioning as adverbial

e.g. 8.28 S(139)
Thinking of that the next hour I have a rest, I spent the whole hour in her office.

In example 8.28, the underlined clause functions as adverbial.

g. Noun phrases: all noun phrases that consist of the head and a determiner and/or a modifier (prenominal or post-nominal) are identified i.e. no noun phrases that consist of the head only are identified.

e.g. 8.29 S(33)
All of this is because of the nature that they live in

e.g. 8.30 S(139)
for me he was a sombol of strength

e.g. 8.31 S(300)
I saw a very beautiful places.

h. Relative clauses: Although relative clauses have already been identified as subordinate or dependent clauses, they have also been identified as post-nominal modifiers.

e.g. 8.32 S(292)
I thanked them for the time they had spent with me.

e.g. 8.33 S(276)
in a country where there is disagreements and depression

e.g. 8.34 S(1)
I went there with my uncle who was youngest than me

i. Other post-nominal modifiers: the other post-nominal modifiers that have been identified are the *to-infinitive phrases*, *participial clauses and phrases* and *prepositional phrases.*

e.g. 8.35 S(82)
that is the best way to develop our country.

e.g. 8.36 S(222)
Every detail known from reality is in those games

e.g. 8.37 S(86)
the other teanagers around him

e.g. 8.38 S(87)
mother tongue's dictionary in your mind

Another point worth noting is that in the case of the occurrence of complex noun phrases where there are relative clauses, prepositional phrases and the noun phrase following the preposition, all their constituents that are embedded in each other are identified. Consider the following example:

e.g. 8.39 S(62)
my friends who showed me the parts of the building and halls of students

This is a complex noun phrase where it can be divided into modifiers as in the following table (8.6)

Table 8.6 the identification of post-nominal modifiers in a sample complex noun phrase

	Sequences	Identified as
1	*my friends who showed me the parts of the building and halls of students*	Noun phrase
2	*who showed me the parts of the building and halls of students*	Relative clause as modifier of the head *friends*
3	*the parts of the building*	another noun phrase
4	*halls of students*	another noun phrase
5	*of the building*	prepositional phrase as a modifier of the head *parts*

6	*the building*	Another noun phrase as the object of the preposition *of*
7	*of students*	prepositional phrase as a modifier of the head *halls)*

8.4 User engagement

20% of the data analyzed for this chapter in terms of complexity has been reanalyzed by a second rater (or as I have called him/her 'a user' because it is more likely that he/she will use the research) in a discussion meeting with me as the first rater. Different points have been discussed among which there were points on which we soon agreed (e.g. the structures that I have missed in rating[4] and brought to my notice by the second rater or the structures that were easy to identify), points that triggered a lot of discussion before reaching an agreement and points that we disagreed on. The following examples exemplify these points of agreement and disagreement.

Examples of the first case

1. e.g. 8.40 S(222) *Young children are not able to make one of their own fantasy.* I have missed rating *to make one of their own fantasy* as a non-finite clause.
2. e.g. 8.41 S(221) *As they are addicted to it, they keep buying it.* The second rater just wondered if *As they are addicted to it* is a dependent finite clause. Then we both agreed that this *as* means *because* and hence the clause is a dependent finite clause.
3. e.g. 8.42 S(292) *Finally, after spending ten days there and returned home again safely.* I have coded this a sentence thinking that it meant *Finally, after spending ten days there I returned home again safely.* The second rater considered this as a fragment and not a sentence.

Examples of the second and third cases:

It is important to note that most of these cases included the decision whether a prepositional phrase is a post-nominal modifier or an adverbial. However, there were cases of dividing the texts into sentences that also sometimes caused some doubt.

The second case

1. e.g. 8.43 S(227) *such kind of communicating between girls and boys.* We had a lot of discussion on whether the prepositional phrase *between girls and boys* is a post-nominal modifier that describes *communication* or it is an adverbial. However, in the end, we decided that it is a post-nominal modifier. The same applies to the example produced by the subject (222) e.g. 8.44 *Most often the first things that come to our mind, when the subject turns towards technology in life, are TV and Internet.* We both discussed whether the phrase *in life* is a post-nominal modifier or adverbial and we agreed that it is a post-nominal modifier that describes the word *technology*.

2. e.g. 8.45 S(228) *First of all, the language will be better if students study All subject in English language, because when the student will finish the school automatically he/she will take in the university and this student will be free and can understand everything and be excellent* (sentence 1). *And do not have a problem with language because for us very difficult when we are in first year, we only have one subject we study in English and we just know the grammar* (sentence 2). We discussed whether *And do not have a problem with language because for us very difficult when we are in first year* goes with sentence 1 or 2. We both reached an agreement that it is a part of sentence 1 and not 2, although orthographically it is part of sentence 2.

Examples of the third case:

1. e.g. 8.46 S(226) *how come!* (sentence 1) *When they were kids they studied in Kurdish language but in bahdini dialect and in high school they study in Surani dialect and at the end in the university or institute they study in Arabic or English, which means there is no Kurdish* (sentence 2). I considered *how come!* as one sentence that is separate from sentence 2 but the second rater suggested that it is not separate from sentence 1.
2. e.g. 8.47 S(228*) I say that english language is a global language in the world and it is a great idea to know this language at the begeaning in the school.* In this sentence, the clause *that english language is a global language in the world* is a dependent clause but is the clause *it is a great idea to know this language at the begeaning in the school* also a dependent clause? What triggered the discussion was the coordinating conjunction *and* between the two clauses: *that english language is a global language in the world* and *it is a great idea to know this language at the begeaning in the school* . The second rater suggested that the

sentence means *I say that english language is a global language in the world and [that] it is a great idea to know this language at the begeaning in the school.* In this case both are dependent clauses. Another example is produced by S(298) e.g. 8.48 *for the first time I saw her at the university of zakho, we were introduced each other.* I considered this as two sentences but the second rater argued that if the phrase *for the first time* means *when*, then it could be considered one sentence.

3. e.g. 8.49 S(229) *That's why people wanted go to the village because in the village have flower tree, plant*. I considered *go to the village* as non-finite although the student has mistakenly not put *to.* The second rater did not consider it a non-finite clause.
4. e.g. 8.50 S (223) *We can see a lot of naturality in living style*. We both (the second rater and I) disagreed on whether *in living style* is a post-nominal modifier or adverbial.

8.5 Essay no. 37 analyzed in detail (Attached as example 8.51 in appendix Q)

First the essay (37) is analyzed into sentences and T-units in table 8.7 and then into noun phrases and post-nominal modifiers together with non-finite clauses and phrases in table 8.8 below

Table 8.7 the division of the essay (37) presented as example 8.51 into sentences, T-units and dependent finite clauses

	T-units	Sentences	Dependent finite clauses
1	*Every people in this world live in different places*	*Every people in this world live in different places and each one of them prefer some places to live in.*	who lived in the past
2	*each one of them prefer some places to live in*	*The old people who lived in the past, they live in countryside because there was no city life on that time.*	because there was no city life on that time
3	*The old people who lived in the past, they live in countryside because there was no city life on that time*	*Nowadays the life become very modern.*	because it's the time of technology and in every step in your life, we will need technological tools and things like internet, mobile, and so on

4	*Nowadays the life become very modern*	*Every individual in this world prefer countryside or City life.*	which we can go and enjoy with our family and friends
5	*Every individual in this world prefer countryside or City life*	*In my point of view, City life is better than the countryside because it's the time of technology and in every step in your life, we will need technological tools and things like internet, mobile, and so on.*	that is not exist in countryside
6	*In my point of view, City life is better than the countryside because it's the time of technology and in every step in your life, we will need technological tools and things like internet, mobile, and so on*	*Also there are many places in city which we can go and enjoy with our family and friends such as parks, cofees, libraries and many other places that is not exist in countryside.*	that create excessive waste and pollution
7	*Also there are many places in city which we can go and enjoy with our family and friends such as parks, cofees, libraries and many other places that is not exist in countryside*	*Of course there are many drawbacks about cities like existence of many factories and places that create excessive waste and pollution.*	which is very bad and unhealthy for us
8	*Of course there are many drawbacks about cities like existence of many factories and places that create excessive waste and pollution*	*This results are the main reason to pollute the enviroment which is very bad and unhealthy for us, but beside this disadvantages of City life, we can not live in countryside because it's became as a habit for us and we can not live without this things that we have in City life.*	because it's became as a habit for us
9	*This results are the main reason to pollute the enviroment which is very bad and unhealthy for us*	*Also life in City is more easier than the countryside because there are many things in City that help us to live easily such as many tools, machines but in*	that we have in City life

		countryside, we have to do every thing by ourselves, and it is very difficult for us because our generation in nowadays are not as strong as the old ones.	
10	*but beside this disadvantages of City life, we can not live in countryside because it's became as a habit for us and we can not live without this things that we have in City life*	*Finally we can say that the city life is a better and easier than the countryside to live in it and also to have a comfortable life.*	because there are many things in City that help us to live easily such as many tools, machines
11	*Also life in City is more easier than the countryside because there are many things in City that help us to live easily such as many tools, machines*		that help us to live easily such as many tools, machines
12	*but in countryside, we have to do every thing by ourselves*		because our generation in nowadays are not as strong as the old ones
13	*and it is very difficult for us because our generation in nowadays are not as strong as the old ones*		that the city life is a better and easier than the countryside to live in it and also to have a comfortable life.
14	*Finally we can say that the city life is a better and easier than the countryside to live in it and also to have a comfortable life.*		

Table 8.8 Noun phrases with their postmodifiers and non-finite clauses as nominals and adverbials identified in the essay number 37 presented as example 8.51 and attached as appendix Q

	NP	PP as PNM	RC as PNM	P as PNM	To-IF as PNM	NFc/ pN	NFc/ pA
1	Every people in this world	in this world					

2	this world						
3	different places						
4	each one of them	of them					
5	some places to live in.				to live in		
6	The old people who lived in the past		who lived in the past				
7	the past						
8	city life						
9	that time						
10	the life						
11	Every individual in this world	In this world					
12	this world						
13	countryside or City life						
14	my point of view	of view					
15	City life						
16	the countryside						
17	the time of technology	of technolog y					
18	every step in your life	in your life					
19	your life						
20	technological tools and things like internet, mobile	like internet, mobile					
21	many places in city which we can go and enjoy with our family and friends such as parks, cofees, libraries and many other places that is not exist in countryside	in City	which we can go and enjoy with our family and friends such as parks, cofees, libraries and many other places that is not exist in countryside				
22	our family and friends						
23	many other places that is not exist in countryside		that is not exist in countryside				
24	many drawbacks about cities like existence of many factories and places that create	(1) about cities (2) like existence of many					

	excessive waste and pollution	factories and places that create excessive waste and pollution					
25	existence of many factories and places that create excessive waste and pollution	of many factories and places that create excessive waste and pollution	that create excessive waste and pollution				
26	excessive waste and pollution						
27	This results						
28	the main reason to pollute the enviroment		which is very bad and unhealthy for us		to pollute the enviro-ment		
29	the enviroment						
30	this disadvantages of City life	of City life					
31	City life						
32	a habit for us	for us					
33	this things that we have in City life.		that we have in City life.				
34	City life.						
35	life in City	in city					
36	the countryside						
37	many things in City that help us to live easily	in City	that help us to live easily				
38	many tools, machines						
39	our generation						
40	the old ones						
41	the city life						
42	the countryside						
43	a comfortable life						

The above listed measures are applied to this essay as follows:

1. *Overall complexity*

- Mean length of T-unit (W/T words per T-unit) = 288÷14= 20.57

This measure shows the average number of words per T-unit

2. *Coordination*

- T-units per sentences (T/S)= 14÷10=1.40

3. *Subordination and embeddedness*

- Dependent and embedded clauses per T-unit (DC/T)= 13÷14=0.93

4. *Phrasal complexity*

- Prepositional phrases as post-nominal modifiers/Total number of noun phrases (PP/NP) = 15÷43=0.35
- Total number of relative clauses / Total number of noun phrases (RC/NP)= 7÷43=0.16
- Total number of *to-infinitive* clauses and phrases as post-nominal modifiers/ Total number of phrases (To-inf./TNP)=2÷43=0.05
- Total number of participial clauses and phrases (*ing-participles and ed-participles*) as post-nominal modifiers/ Total number of noun phrases (P/NP)=0÷43= 0

5. *Clausal complexity*

Measures of infinitive clauses and phrases functioning as nominal and ing and ed-clauses and phrases functioning as nominals and adverbials

- Non-finite clauses/phrases as nominals and adverbials / total number of T-units (NF Clauses, Phrases/T)

= 0÷14=0

Table 8.9 the complexity index of the essay 37

W/T	T/S	DC/T	PP/NP	RC/NP	P□phases/NP	IF/NP	NF□Clauses,□ Phrases/T
20.57	1.40	0.93	0.35	0.16	0.00	0.05	0.00

This table 8.9 shows the sentence, clause and phrasal complexity of essay 37. The figure in the first column (20.57) is the mean length of T-unit (word per T-unit); the figure in the second column (1.40) is the ratio of the T-units to sentences; the figure in the third column (0.93) is the ratio of dependent clauses to T-unit; the figure in the fourth column (0.35) is the ratio of prepositional phrases as post-nominal modifiers to the total number of noun phrases in the essay; the figure in the fifth column (0.16) is the ratio of relative clauses to the total number of noun phrases in the essay; the figure in the sixth column (0.00) is the ratio of participial phrases as post-nominal modifiers in noun phrases to the total number of noun phrases; the figure in the seventh column (0.05) is the ratio of the *to-infinitive* phrases as

post-nominal modifiers to the total number of noun phrases in the essay; the figure in the eighth column (0.00) is the ratio of non-finite nominal and adverbial *to-infinitive* and ing-clauses and phrases to the total number of T-units.

In this essay, the average length of T-unit is 20.57 per T-unit (the overall complexity of the essay). As for the sentential complexity in terms of coordination and subordination/embeddedness index, this student has relied on coordination by an average of 1.4 coordinated T-units per sentence and subordination and embeddedness by an average of 0.93 dependent and embedded clauses per T-unit. This shows that on average not all sentences contain coordination because there is mostly one T-unit per sentence, but the student has used subordination in most of his/her T-units and this is shown by the figure 0.93 which is close to 1. Still considering sentential complexity, this student has used no non-finite nominal and adverbial *to-infinitive* and ing-clauses and phrases per T-unit. As for the noun phrase complexity as far as postmodification is concerned, each noun phrase contains the average of 0.35 prepositional phrases and 0.16 relative clauses, meaning that only few noun phrases have prepositional phrases and even fewer have relative clauses as post-nominal modifiers. If these two figures were both equal to 1, that would mean that every noun phrase had on average one prepositional phrase and one relative clause as a post-nominal modifier.

8.6 Results and discussion

8.6.1 Accuracy compared to the complexity of the same group of narrative essays

In this section the accuracy scores obtained in chapter five for a number of essays are compared to the complexity scores obtained for the same essays in this chapter

Figure 8.2 the ratio of error-free T-units and error-free clauses for 10 3rd year students/ school of Arts

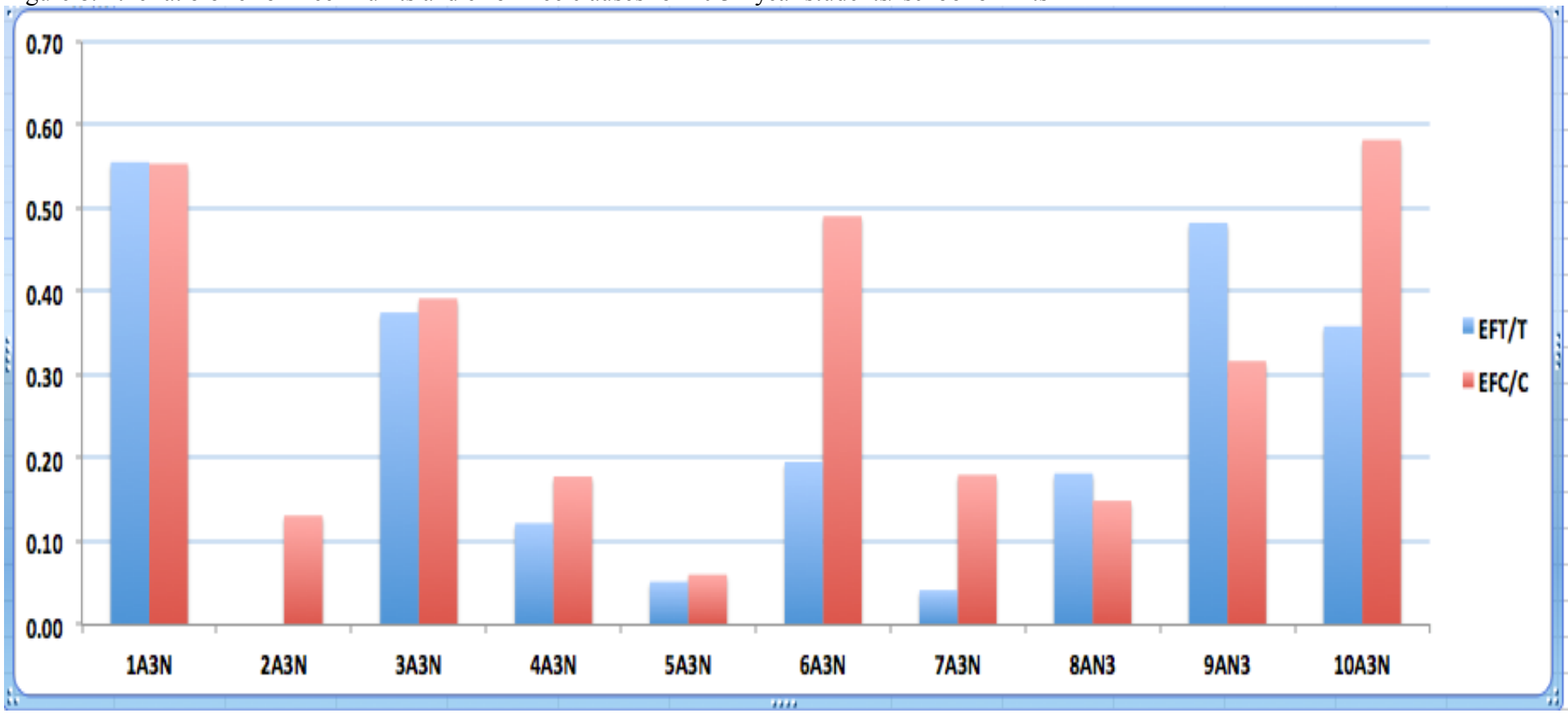


Figure 8.2 shows the ratio of error-free T-units to the total number of T-units (EFT/T) and error-free clauses to the total number of clauses (EFC/C) for 10 third year students of the school of Arts. The horizontal axis shows the students' (or essays) codes; for example, 1A3N is essay number 1 written by a student from the school of Arts (A) in third year of study (3) and its type is narrative (N). The vertical axis represents the ratios stated above.

Figure 8.3 the words per T-unit (the average number of words per T-unit) for 10 3rd year students/ school of Arts

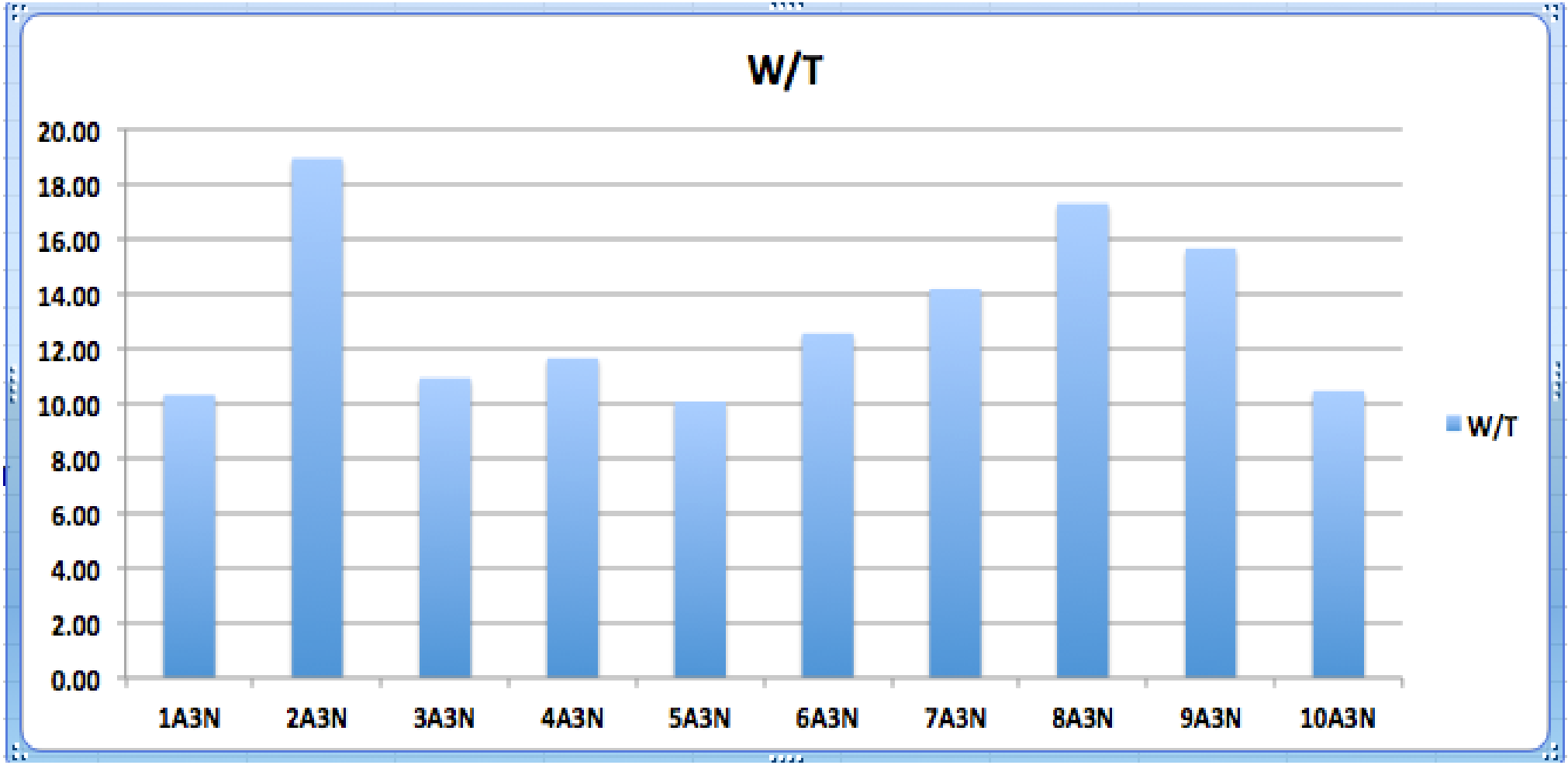


Figure 8.3 shows the average number of words per T-units i.e. how many words there are on average in a T-unit or the mean length of T-unit for 10 third year students from the school of Arts. The horizontal axis shows the students' codes and the vertical axis shows the ratios.

Figure 8.4 the ratio of T-unit per sentence (the coordination index) for 10 3rd year students/ school of Arts

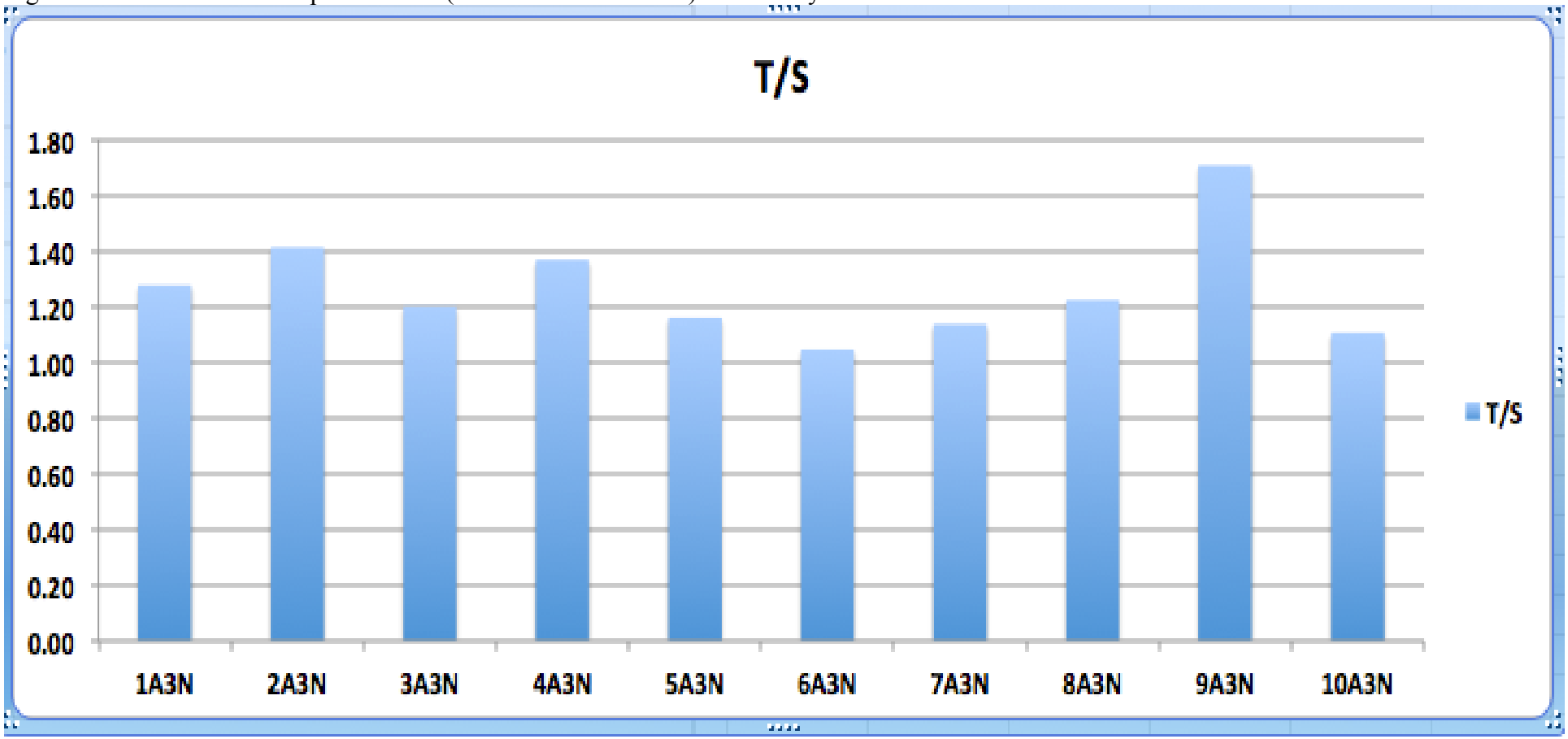
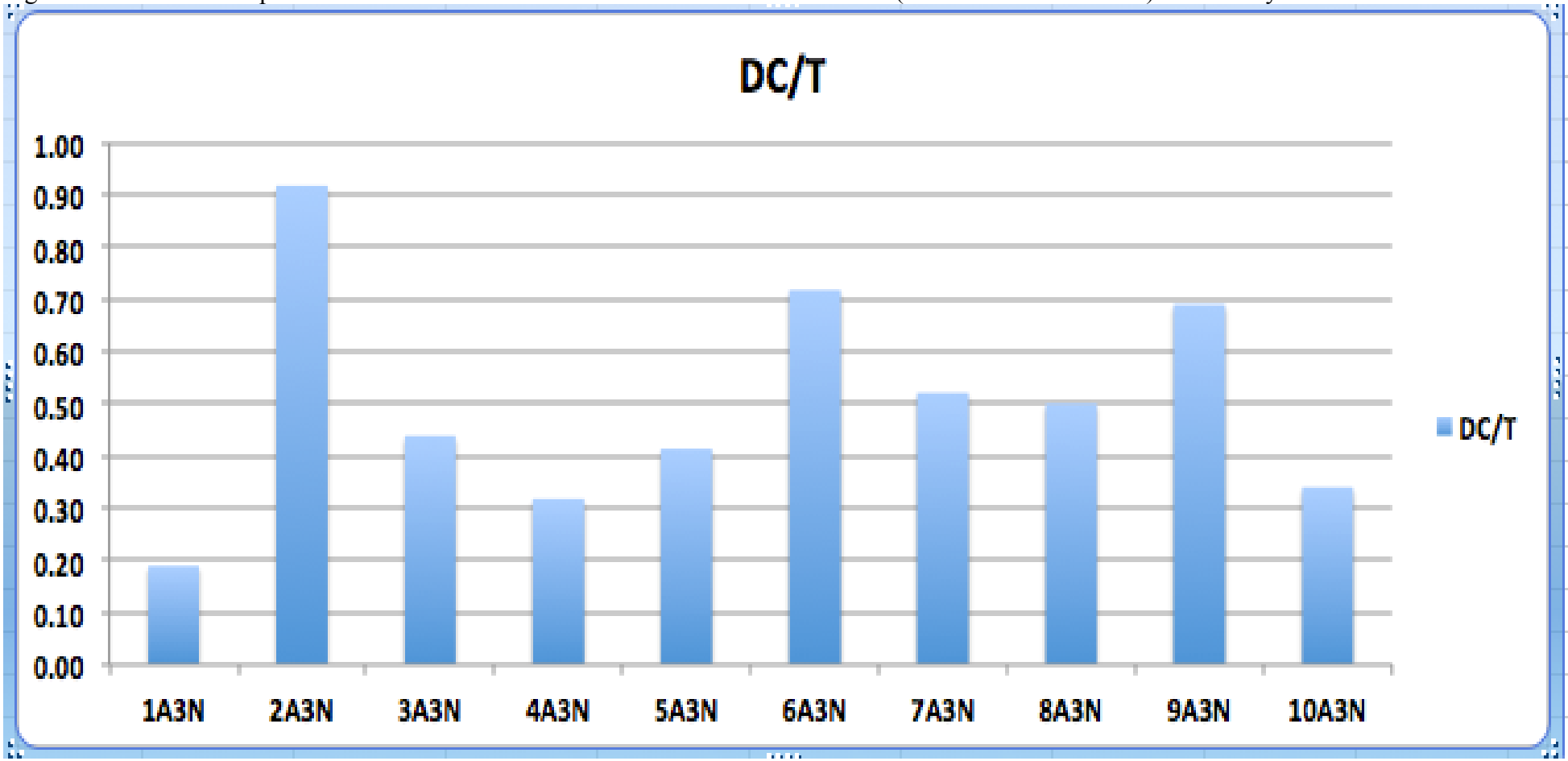


Figure 8.4 shows the ratio of T-units in a sentence, which is a measure of the number of coordinated T-units within one sentence for 10 third year students from the school of Arts. The horizontal axis shows the students' codes and the vertical axis shows the ratios.

Figure 8.5 the ratio of dependent and embedded clauses to the total number of T-units (the subordination index) for 10 3rd year students/ school of Arts



This figure (8.5) shows the ratio of dependent and embedded clauses in each T-unit for 10 third year students from the school of Arts. The horizontal axis shows the students' codes and the vertical axis shows the ratios.

Figure 8.6 the ratio of relative clauses to the total number of noun phrases for 10 3rd year students/ school of Arts

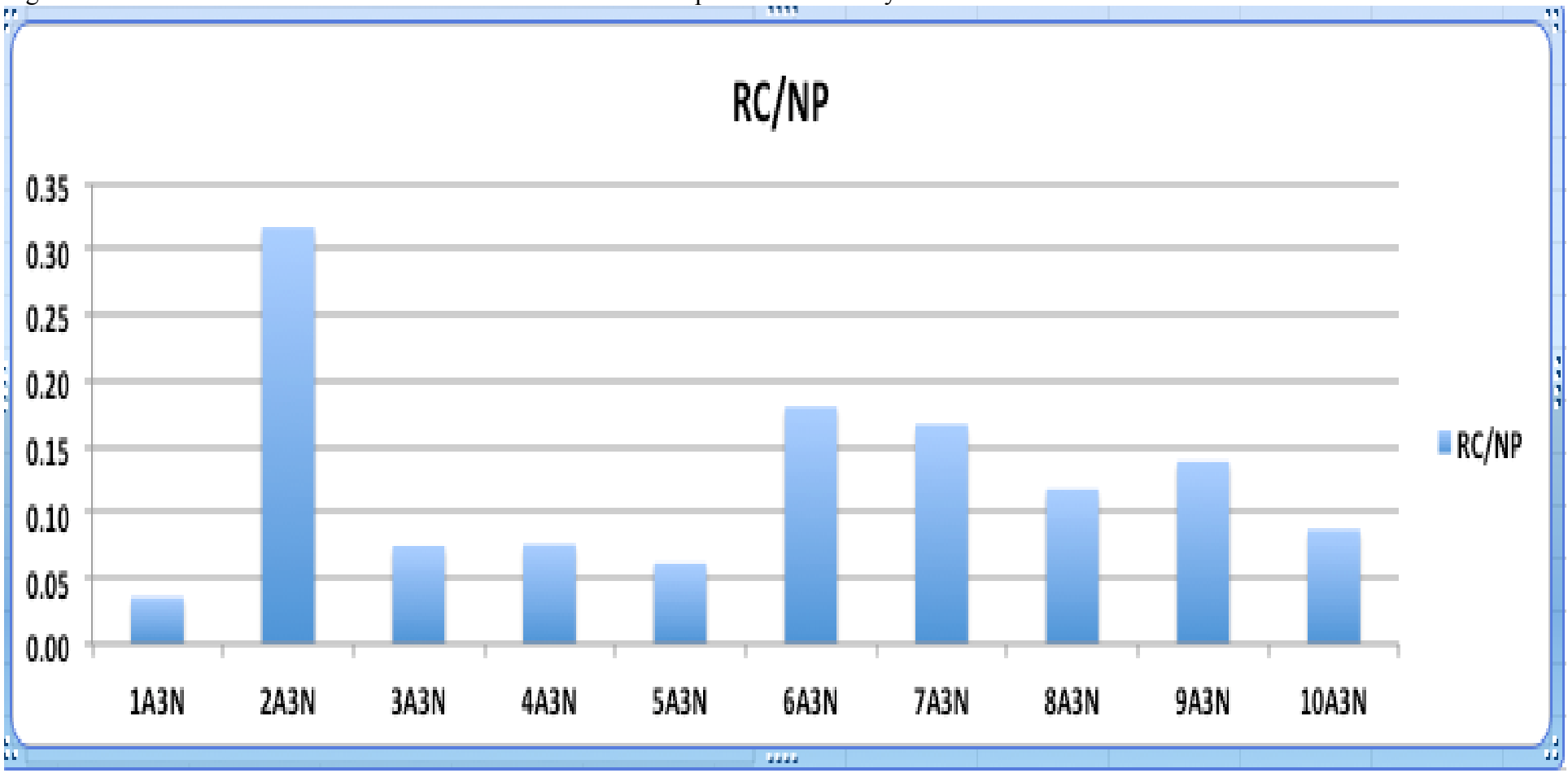


Figure 8.6 shows the ratio of all relative clauses to the total number of noun phrases or the average number of relative clauses per noun phrase for 10 third year students from the school of Arts. The horizontal axis shows the students' codes and the vertical axis shows the ratios.

Figure 8.7 the ratio of prepositional phrases as post-nominal modifiers to the total number of noun phrases for 10 3rd year students/ school of Arts

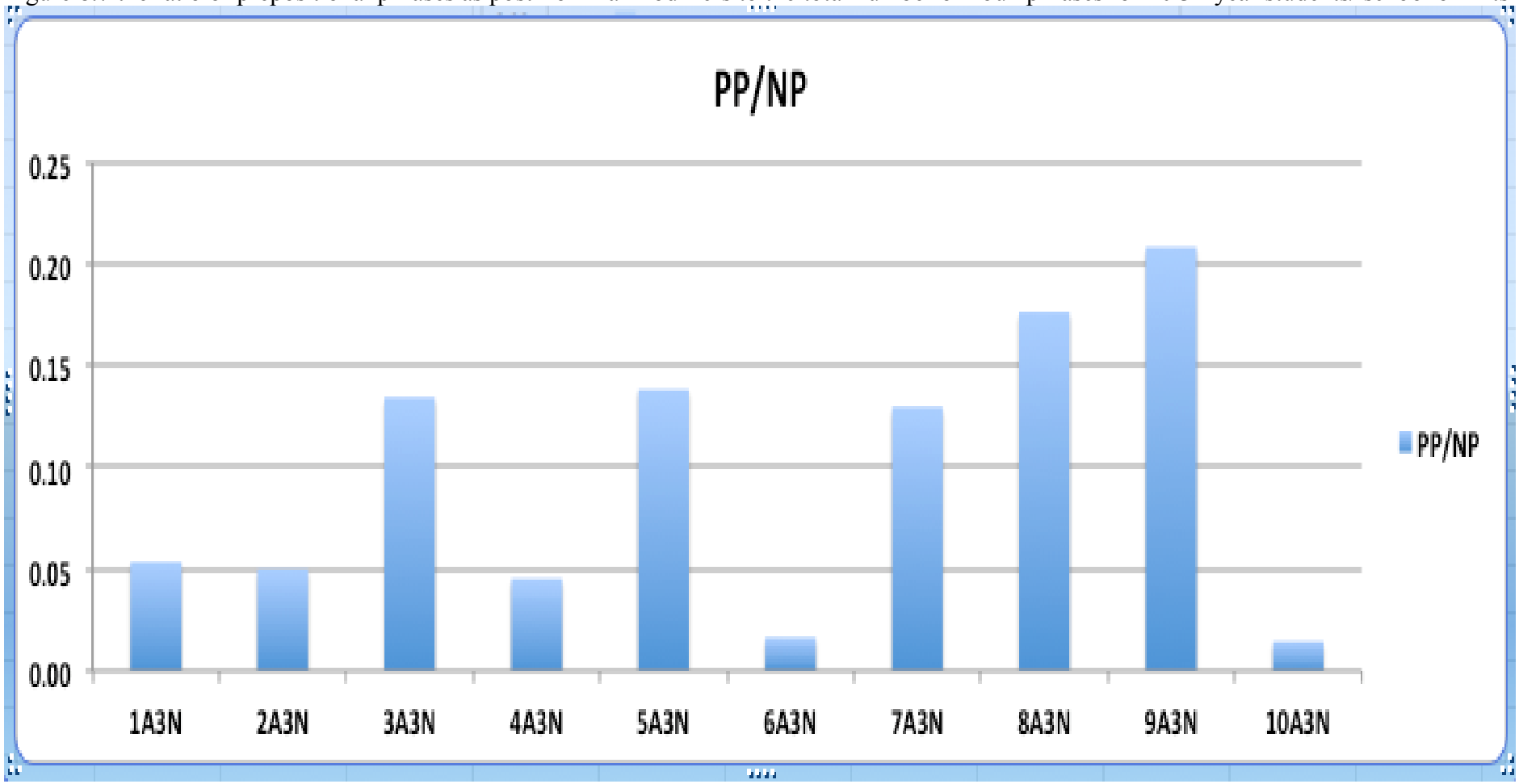


Figure 8.7 shows the ratio of prepositional phrases as post-nominal modifiers to the total number of noun phrases for 10 third year students from the school of Arts. The horizontal axis shows the students' codes and the vertical axis shows the ratios.

Figure 8.8 the ratio of *to-infinitive* phrases as post-nominal modifiers to the total number of noun phrases for 10 3rd year students/ school of Arts

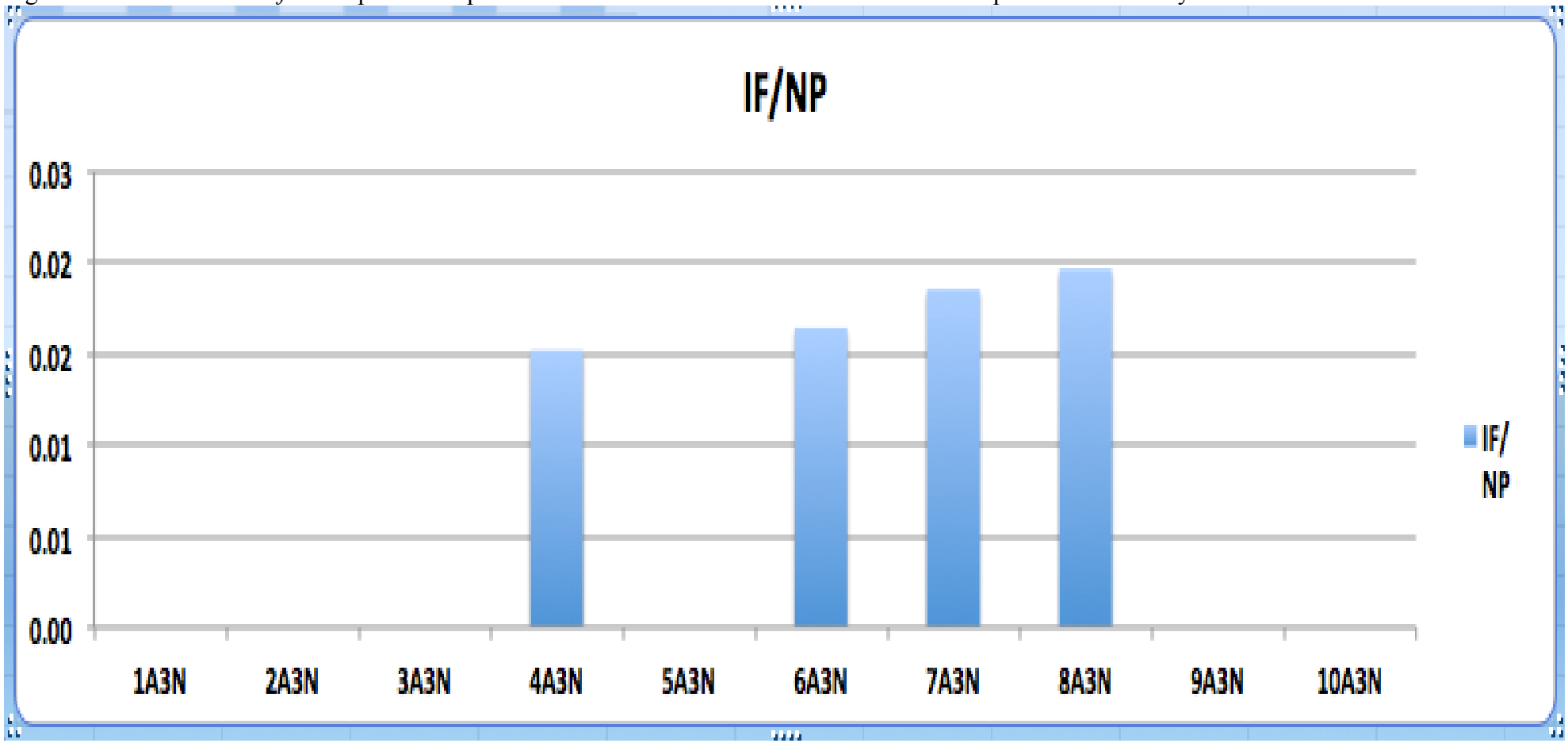


Figure 8.8 shows the ratio of *to-infinitive* phrases as post-nominal modifiers to the total number of noun phrases for 10 third year students from the school of Arts. The horizontal axis shows the students' codes and the vertical axis shows the ratios.

Figure 8.9 the ratio of participial clauses and phrases as post-nominal modifiers to the total number of noun phrases for 10 3rd year students/ school of Arts

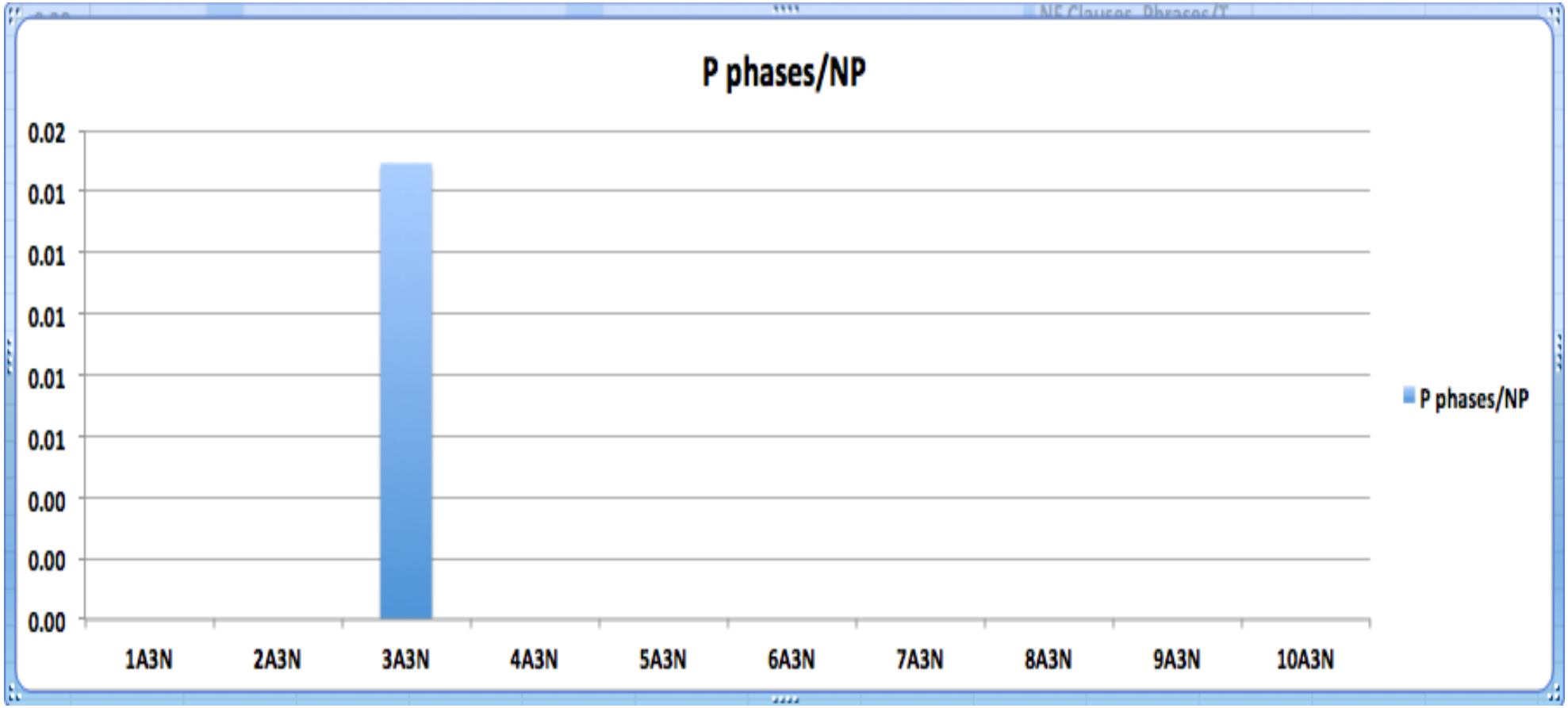


Figure 8.9 shows the ratio of participial phrases and clauses as post-nominal modifiers to the total number of noun phrases for 10 third year students from the school of Arts. The horizontal axis shows the students' codes and the vertical axis shows the ratios.

Figure 8.10 the ratio of non-finite clauses and phrases per /T-unit for 10 3rd year students/ school of Arts

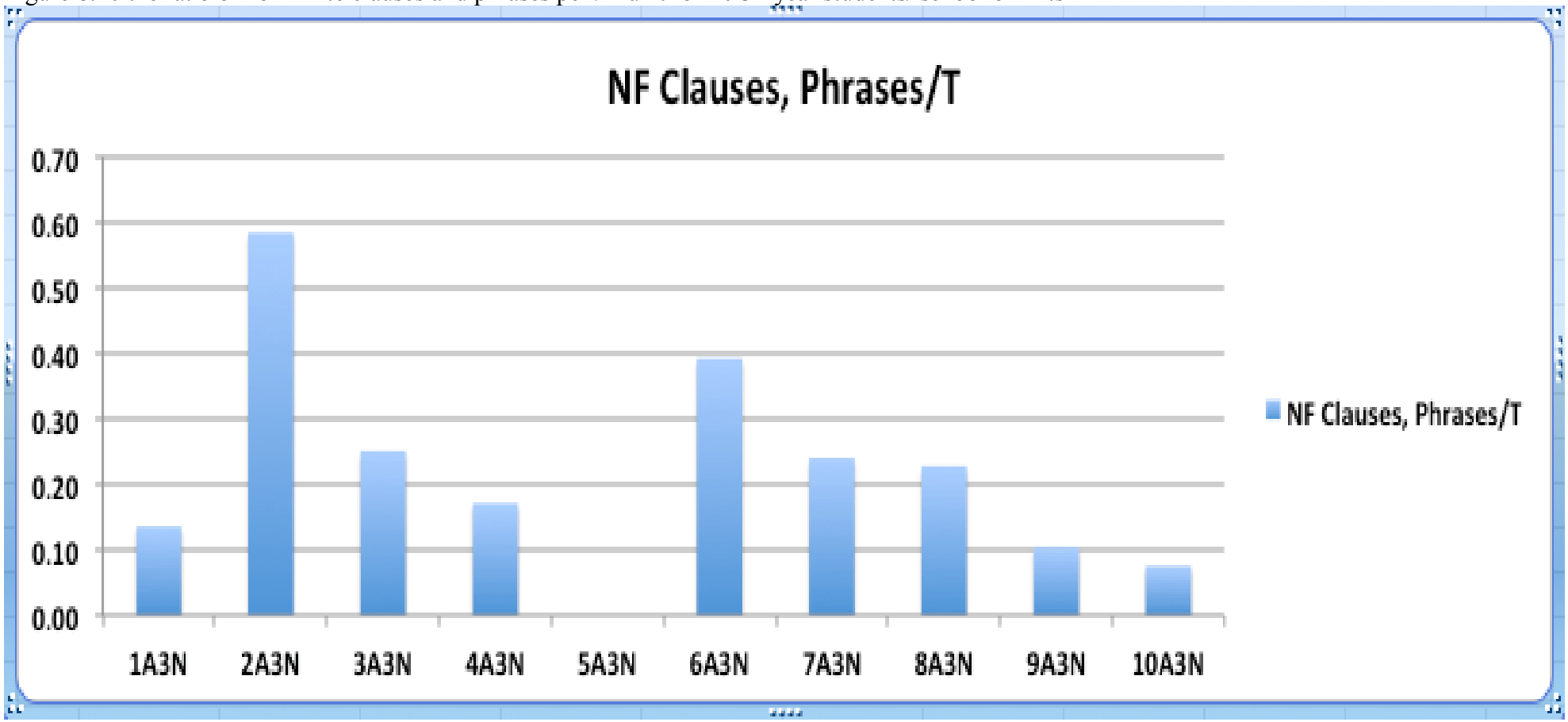


Figure 8.10 shows the ratio of non-finite clauses and phrases to the total number of T-units for 10 third year students from the school of Arts. The horizontal axis shows the students' codes and the vertical axis shows the ratios.

These figures indicate that a number of students could not achieve a balance between complexity and accuracy. In other words, good performance in one of them was at the expense of the other. Students 1A3N, 2A3N and 10A3N can be given as examples. Looking at the figures 8.2 to 8.10, one can see that the students 1A3N and 10A3N have done well on accuracy but they have produced fairly short T-units with little subordination and few complex noun phrases. This is clear from the low ratio of dependent clauses to the total number of the T-units and the low ratio of prepositional phrases, and relative clauses to the total number of noun phrases in the essay. The low phrasal complexity is also obvious from the absence of any *to-infinitive* phrases in the noun phrases, and the low clausal and phrasal complexity is obvious from the very low use of the non-finite clauses and phrases in the nominal and adverbial position. These students seem to have relied more on coordination than subordination because the ratio of T-units in one sentence is higher than the ratio of dependent clauses to the total number of T-units.

Student 2A3N can be considered the opposite of student 1A3N and 10A3N. This student has scored very low on accuracy but has produced very long T-units with a great degree of subordination and highly complex noun phrases with a good number of relative clauses as post-nominal modifiers. The high ratio of non-finite clauses in nominal and adverbial position is also another indicator of the complexity of this student's language. Student 3A3N could be regarded another example of not keeping a balance between complexity and accuracy if compared to the students 1A3N and 10A3N. This student, compared to student 1A3N and 10A3N has scored lower on accuracy but has produced longer T-units, more complex noun phrases and more subordination than student 10A3N. However, both students 1A3N and 3A3N have scored almost the same on coordination.

Other students seem to have avoided a trade-off between accuracy and complexity. Students 6A3N and 9A3N are good examples of keeping this balance. They have produced language that is both accurate and fairly complex. Their scores on figures 8.2 to 8.10 show their phrasal complexity (except for the score of student 6A3N on the ratio PP/NP, and the scores of student 9A3N on the ratio IF/NP). Also both these students, like all the others, have used zero participial phrases and clauses as post-nominal modifiers.

The other students (4A3N, 5A3N, 7A3N, 8A3N) have performed less well on accuracy. They, however, have produced moderately complex language in terms of phrases, T-units and coordinated sentences with students 7A3N, 8A3N being better than the other two i.e. 4A3N, 5A3N. It is significant, though, to mention that coordination might not necessarily be a sign

of improved or even more complex language but rather avoidance of writing complex sentences or the use of simple sentences coordinated thinking that this makes the language more complex. Sometimes the use of complex sentences with a heavy use of subordinating conjunctions like *because* might also not be a good indicator of improvement or even complexity. On the contrary, this may show that the learner has little information about how to make the language more complex by using other devices of complexity like the use of non-finite clauses and phrases or the use of complex noun phrases and hence resorts to the use of frequently used complexity devices such as *because*. Consider the following examples:

e.g. 8.52 S(299)

At the end I confess it was a really hard times for me but also I spent a great time in college and studying, reading time was hard but I learned a lot of things and I took alot of benefits from those and colleges all things that I can say about those days are these and I can't never forget my college in rest of my life and I realize that I will wish for returning back those days.

e.g. 8.53 S(33)

Because I think that the water in the cities is more fresh than water in the country side because now adays there are so many projects in the cities that fresh or clean the water from any(bacterias)because somany people dies be cause a (bacterias)that appear in the nature that can not be destroyed in the country sides

These examples show that not all types of complexity are a sign of improvement. Some kinds, such as the excessive use of coordination or the use of a string of subordinated clauses connected with one subordinate conjunction like *because* might indicate that learners have not developed in their writing yet because "developing writers use *because* less frequently as their writing matures and they learn the conventions of academic written English" (Crowhurst 1987 cited in Schleppegrell, 1996:273) and hence they compensate for their lack of awareness of the use of correct punctuation and other devices of complexity by using such long sentences or T-units. It is also important to emphasize that subordination with different subordinating conjunctions might not be with the same linguistic complexity. *Because* is one of the frequently used subordination conjunctions and university ESL writers make an excessive use of *because* (Schleppegrell, 1996) but an item like *since* seems to be less available than *because*. If the measures of subordination and coordination that I have used in this research consisted of measures of the devices with which this subordination and coordination are formed, the findings might have been different. For example, if the subordination with *because* was measured separately from the subordination with other subordinating

conjunctions, the picture might have been different as it would have let the reader know that *because* is more common than the other subordinating conjunctions if the case was so.

Lambert and Kormos (2014:608) have brought to attention an argument similar to this. They have also argued against operationalizing subordination as "a unitary construct" because this method of dealing with subordination ignores the different types of subordination and subordination with different devices or items. In other words, it is important to note that measures of complexity have to be treated with caution, and that we need to see each measure separately rather than applying them together as measures of one construct (e.g. subordination).

Before concluding this section a word is in order about the participial phrases in figure 8.9. It is clear that only one student has produced one instance of this structure. This could point to the fact that such participial clauses, although reduced, are linguistically complex as they involve two processes i.e. relativization and reduction (see example 8.54) which might also make them cognitively difficult for learners to produce. This could make us think that these structures might be the most complex ones linguistically as well. This shows that for a structure to be linguistically complex, it does not have to be long.

e.g. 8.54 *I greeted the girl . She is standing in the corridor*
Process (1) Relativization: *I greeted the girl who is standing in the corridor*
Process (2) Reduction: *I greeted the girl standing in the corridor*.

I have not found even one instance of these ing-clauses in all the sample analyzed and only two instances of ed-clauses as post-nominal modifiers

8.6.2 Comparison of the complexity scores of narrative and argumentative essays between years

Table 8.10 the complexity scores of the narrative essays of 10 third year students at the school of Arts.

Students□codes	W/T	T/S	DC/T	PP/NP	RC/NP	P□phases/NP	IF/NP	NF□Clauses,□ Phrases/T
1A3N	10.30	1.28	0.19	0.05	0.04	0.00	0.00	0.14
2A3N	18.92	1.41	0.92	0.05	0.32	0.00	0.00	0.58
3A3N	10.92	1.20	0.44	0.13	0.07	0.01	0.00	0.25
4A3N	11.63	1.37	0.32	0.05	0.08	0.00	0.02	0.17
5A3N	10.07	1.16	0.41	0.14	0.06	0.00	0.00	0.00
6A3N	12.54	1.05	0.72	0.02	0.18	0.00	0.02	0.39
7A3N	14.16	1.14	0.52	0.13	0.17	0.00	0.02	0.24
8A3N	17.27	1.22	0.50	0.18	0.12	0.00	0.02	0.23
9A3N	15.62	1.71	0.69	0.21	0.14	0.00	0.00	0.10
10A3N	10.45	1.10	0.34	0.01	0.09	0.00	0.00	0.08

Table 8.10 tabulates the complexity scores of 10 narrative essays written by 10 third year students at the school of Arts. The figures in the first column are the mean length of T-unit (word per T-unit); the figures in the second column are the ratio of the T-units to sentences; the figures in the third column are ratio of dependent clauses to T-unit; the figures in the fourth column are the ratio of prepositional phrases as post-nominal modifiers to the total number of noun phrases in the essay; the figures in the fifth column are the ratio of relative clauses to the total number of noun phrases in the essay; the figures in the sixth column are the ratio of participial phrases as post-nominal modifiers in noun phrases to the total number of noun phrases; the figures in the seventh column are the ratio of the *to-infinitive* phrases as post-nominal modifiers to the total number of noun phrases in the essay; the figures in the eighth column are the ratio of non-finite nominal and adverbial *to-infinitive* and ing-clauses clauses and phrases to the total number of T-units.

Table 8.11 the complexity scores of the argumentative essays of 10 third year students at the school of Arts

Students□codes	W/T	T/S	DC/T	PP/NP	RC/NP	P□phases/NP	IF/NP	NF□Clauses,□ Phrases/T
33A3A	16.18	1.47	0.68	0.24	0.10	0.00	0.00	0.09
34A3A	12.77	1.05	0.33	0.36	0.01	0.00	0.00	0.03
35A3A	20.05	1.33	0.55	0.10	0.02	0.00	0.00	0.25
36A3A	14.83	1.14	0.64	0.08	0.14	0.00	0.06	0.07
37A3A	20.57	1.40	0.93	0.37	0.16	0.00	0.05	0.00
38A3A	36.55	1.10	2.45	0.14	0.17	0.00	0.00	0.91
39A3A	13.80	1.92	0.40	0.21	0.06	0.00	0.00	0.00
40A3A	9.51	1.18	0.16	0.16	0.02	0.00	0.00	0.09
41A3A	11.86	1.06	0.35	0.19	0.15	0.00	0.04	0.03
42A3A	19.31	1.38	0.69	0.14	0.05	0.00	0.00	0.14

Table 8.11 tabulates the complexity scores of 10 argumentative essays written by 10 third year students at the school of Arts. W/T=total number of words divided by the total number of T-units; T/S=total number of T-units divided by the total number of sentences; DC/T=total number of dependent and embedded clauses divided by the total number of T-units; PP/NP=total number of prepositional phrases as post-nominal modifiers divided by the total number of noun phrases; RC/NP=total number of relative clauses divided by the total number of noun phrases; P phrases/NP=total number of participial phrases as post-nominal modifiers divided by the total number of noun phrases; IF/NP=total number of *to-infinitive* phrases divided by the total number of noun phrases; NF clauses, phrases/T=the total number of non-finite clauses and phrases divided by the total number of T-units.

I will first start with the points of similarity between the two tables. The students are very similar in their production of the participial phrases and *to-infinitive* phrases as post-nominal modifiers. They produced hardly any such structures or did not produce any. The students have, nonetheless, been different in their production of other structures. For example, the mean length of the T-unit is considerably higher in the case of argumentative essays (ranging between 36.55 and 9.51 with the figures in between being very high) compared to that of the T-unit in narrative essays (ranging between 18.92 and 10.07).

As for the ratio of T-units to sentences or the coordination index, the students who have written argumentative essays seem to have relied more on coordination (although not to a great extent) than the students who wrote narrative essays as their scores range between 1.05 and 1.92 with the figures in between also being high. The same applies (again not with a remarkable difference) to the ratio of dependent and embedded clauses to the total number of T-units.

As for phrasal complexity, the argumentative essays are more complex in terms of prepositional phrases per noun phrases but not in terms of relative clauses per noun phrases. However, it is noteworthy to highlight that the ratio of non-finite clauses and phrases is higher with the narrative essays, although the essay that scored the highest is argumentative (0.91). This discussion tells us that two aspects of complexity are higher with the narrative essays i.e. the relative clauses and the non-finite clauses and phrases but the argumentative essays have scored higher in complexity as far as mean length of T-units, T-units per sentence, dependent clauses per T-unit and prepositional phrases per noun phrases are concerned. This indicates that the argumentative essays are more complex especially with subordination and coordination though these two aspects might not be good indicators of high complexity because, as Biber *et al.* (2011) highlight, phrasal complexity is one of the features of academic writing rather than clausal and sentential complexity.

Table 8.12 the complexity scores of the narrative essays of 10 fourth year students at the school of Arts

Students codes	W/T	T/S	DC/T	PP/NP	RC/NP	P phases/NP	IF/NP	NF Clauses, Phrases/T
58A4N	16.12	1.55	0.41	0.08	0.16	0.00	0.00	0.18
59A4N	14.00	1.32	0.39	0.20	0.05	0.00	0.04	0.24
60A4N	13.30	1.76	0.30	0.28	0.09	0.00	0.00	0.24
61A4N	13.52	1.48	0.42	0.25	0.02	0.00	0.00	0.13
62A4N	19.00	1.28	1.16	0.31	0.18	0.00	0.00	0.34
63A4N	11.05	1.65	0.42	0.19	0.11	0.00	0.00	0.19
64A4N	11.21	1.26	0.38	0.16	0.05	0.00	0.00	0.10
65A4N	17.80	1.59	0.43	0.09	0.07	0.00	0.00	0.40
66A4N	15.79	1.32	0.76	0.21	0.19	0.00	0.00	0.12
67A4N	12.02	1.31	0.41	0.16	0.05	0.00	0.00	0.12

Table 8.12 tabulates the complexity scores of 10 narrative essays written by 10 fourth year students at the school of Arts.

Table 8.13 the complexity scores of the argumentative essays of 10 fourth year students at the school of Arts

Students codes	W/T	T/S	DC/T	PP/NP	RC/NP	P phases/NP	IF/NP	NF Clauses, Phrases/T
81A4A	11.64	1.04	0.36	0.30	0.00	0.00	0.01	0.19
84A4A	20.70	1.03	0.47	0.30	0.08	0.00	0.00	0.97
85A4A	11.98	1.53	0.29	0.18	0.05	0.00	0.01	0.12
86A4A	31.67	1.33	1.08	0.19	0.08	0.00	0.00	0.50
87A4A	22.74	1.53	0.91	0.15	0.07	0.00	0.02	0.22
88A4A	18.64	1.25	0.84	0.16	0.13	0.00	0.00	0.20
89A4A	17.33	1.20	1.08	0.18	0.06	0.00	0.00	0.42
90A4A	16.89	1.27	0.68	0.23	0.14	0.00	0.00	0.32
92A4A	12.64	1.57	0.32	0.31	0.11	0.00	0.00	0.05
93A4A	13.16	1.58	0.47	0.12	0.06	0.00	0.00	0.00

Table 8.13 tabulates the complexity scores of 10 argumentative essays written by 10 fourth year students at the school of Arts.

To compare tables 8.12 and 8.13, it is again obvious that there are very few, if any, participial and *to-infinitive* phrases in both tables. T-units are also longer in the argumentative essays than those in the narrative essays with an average range 31.67–11.64 compared to an average range 19.00 –11.05. Unlike tables 8.10 and 8.11, the ratio of T-units to sentences, the coordination index, is higher, in the case of narrative essays than in the case of argumentative essays. Concerning the subordination index or the ratio of dependent clauses to T-units, there

is not very big difference between the two types of essays though the argumentative essays seem to have scored higher. Also, in the case of phrasal complexity, the ratio of prepositional phrase to the noun phrase appears to be higher in the argumentative essays with an average range between 0.31–0.12 compared to an average range between 0.31–0.08. Concerning the complexity of noun phrase in terms of the ratio of relative clauses, the narrative essays have more relative clauses than the argumentative essays. The ratio of non-finite clauses and phrases per T-unit is higher in the case of argumentative essays.

Table 8.14 the complexity scores of the narrative essays of 10 third year students at the school of Basic Education

Students codes	W/T	T/S	DC/T	PP/NP	RC/NP	P phases/NP	IF/NP	NF Clauses, Phrases/T
113B3N	18.11	1.06	0.95	0.29	0.12	0.00	0.03	0.16
114B3N	11.73	1.38	0.36	0.06	0.29	0.00	0.00	0.05
115B3N	11.29	2.50	0.43	0.25	0.04	0.00	0.00	0.03
116B3N	13.45	1.32	0.48	0.12	0.12	0.00	0.00	0.21
117B3N	8.78	1.19	0.22	0.25	0.04	0.00	0.00	0.02
118B3N	11.55	1.07	0.52	0.10	0.05	0.00	0.00	0.17
119B3N	16.13	1.07	0.69	0.11	0.14	0.00	0.03	0.00
120B3N	8.85	1.62	0.32	0.06	0.09	0.00	0.00	0.03
121B3N	13.44	1.30	0.21	0.04	0.10	0.00	0.01	0.19
122B3N	13.73	1.25	0.64	0.26	0.08	0.00	0.00	0.29

Table 8.14 tabulates the complexity scores of 10 narrative essays written by 10 third year students at the school of Basic Education.

Table 8.15 the complexity scores of the argumentative essays of 10 third year students at the school of Basic Education

Students codes	W/T	T/S	DC/T	PP/NP	RC/NP	P phases/NP	IF/NP	NF Clauses, Phrases/T
191B3A	19.46	1.04	0.88	0.09	0.06	0.00	0.04	0.13
192B3A	16.80	1.15	0.53	0.15	0.05	0.00	0.02	0.27
193B3A	13.71	1.23	0.55	0.10	0.06	0.00	0.02	0.21
195B3A	16.35	2.06	0.57	0.15	0.06	0.00	0.01	0.11
196B3A	18.57	1.00	0.57	0.25	0.12	0.00	0.00	0.33
197B3A	25.97	1.20	0.73	0.10	0.10	0.00	0.03	0.00
198B3A	13.81	1.04	0.77	0.11	0.13	0.00	0.04	0.42
199B3A	19.85	1.23	0.74	0.04	0.01	0.00	0.01	0.11
200B3A	23.95	1.11	1.00	0.20	0.13	0.00	0.01	0.30
201B3A	12.13	1.07	0.34	0.10	0.05	0.00	0.00	0.11

Table 8.15 tabulates the complexity scores of 10 argumentative essays written by 10 third year students at the school of Basic Education.

Like the case with the tables above, the mean length of T-units in the argumentative essays is higher than in the narrative essays and the participial phrases are not produced in both types. However, the number of *to-infinitive* phrases as post-nominal modifiers is higher in the case of argumentative essays while the coordination index is higher in the case of narrative essays. Concerning the ratio of dependent and embedded clauses, the argumentative essays are obviously more complex. In phrasal complexity, the ratio of prepositional phrases and relative clauses is not very different with both types of essays but a little higher in narrative essays than in the argumentative essays. The ratio of non-finite clauses and phrases (*to-infinitive*) in nominal and the ratio of ing-clauses and phrases in nominal and adverbial position is higher in the argumentative essays than in the narrative essays. To sum up, the argumentative essays seem to be more complex in terms of clausal complexity but the narrative essays seem to be a little more complex at the level of phrasal complexity.

Table 8.16 the complexity scores of the narrative essays of 10 fourth year students at the school of Basic Education

Students codes	W/T	T/S	DC/T	PP/NP	RC/NP	P phases/NP	IF/NP	NF Clauses, Phrases/T
139B4N	12.28	1.59	0.51	0.12	0.09	0.00	0.02	0.05
140B4N	15.06	1.07	0.53	0.15	0.08	0.00	0.02	0.34
141B4N	15.03	1.43	0.47	0.14	0.07	0.00	0.00	0.20
142B4N	8.50	1.00	0.38	0.22	0.00	0.00	0.00	0.16
143B4N	15.27	1.43	0.70	0.23	0.14	0.00	0.00	0.13
144B4N	15.68	1.32	0.52	0.27	0.02	0.00	0.00	0.04
145B4N	11.82	1.21	0.35	0.04	0.13	0.08	0.00	0.18
146B4N	12.85	1.11	0.40	0.21	0.05	0.00	0.00	0.00
147B4N	14.91	1.53	0.48	0.07	0.10	0.00	0.00	0.13
148B4N	8.19	1.28	0.05	0.05	0.04	0.00	0.00	0.07

Table 8.16 tabulates the complexity scores of 10 narrative essays written by 10 fourth year students at the school of Basic Education.

Table 8.17 the complexity scores of the argumentative essays of 10 fourth year students at the school of Basic Education

Students codes	W/T	T/S	DC/T	PP/NP	RC/NP	P phases/NP	IF/NP	NF Clauses, Phrases/T
161B4A	12.64	1.08	0.43	0.14	0.14	0.00	0.00	0.25
162B4A	18.60	1.25	0.80	0.24	0.12	0.00	0.00	0.53
167B4A	11.10	1.35	0.29	0.14	0.07	0.00	0.00	0.13
168B4A	15.22	1.23	0.41	0.13	0.03	0.00	0.02	0.11
169B4A	11.55	1.27	0.47	0.12	0.04	0.00	0.00	0.16
171B4A	20.56	1.13	1.28	0.23	0.16	0.00	0.00	0.56
172B4A	12.64	1.12	0.21	0.30	0.03	0.00	0.00	0.18
173B4A	12.13	1.14	0.29	0.24	0.09	0.00	0.04	0.13
174B4A	14.12	1.14	0.60	0.00	0.17	0.00	0.00	0.28
175B4A	14.54	1.04	0.57	0.11	0.07	0.00	0.02	0.14

Table 8.17 tabulates the complexity scores of 10 argumentative essays written by 10 fourth year students at the school of Basic Education.

To compare table 8.16 to table 8.17, the similarity lies more in the ratio of participial phrases with only one narrative essay having a few participial phrases. Also, the ratio of *to-infinitive* phrases as post-nominal modifiers is very much the same except for a difference of one essay. Like all the previous cases, the mean length of the T-unit is higher in the case of argumentative essays but with a slight difference (not like the other cases). The ratio of T-units to sentences is higher in the case of narrative essays; the subordination index does not seem to be very different although it reached higher levels in argumentative essays; the phrasal index in terms of both prepositional phrases and relative clauses as post-nominal modifiers is higher in the case of argumentative essay as well. However, the non-finite clauses ratio is noticeably higher in the argumentative essays.

Table 8.18 the complexity scores of the narrative essays of 10 third year students at the school of Languages

Students codes	W/T	T/S	DC/T	PP/NP	RC/NP	P phases/NP	IF/NP	NF Clauses, Phrases/T
250L3N	12.74	2.88	0.30	0.21	0.03	0.00	0.03	0.04
252L3N	12.97	1.30	0.57	0.10	0.15	0.01	0.01	0.14
253L3N	8.73	1.94	0.12	0.13	0.00	0.03	0.03	0.18
254L3N	12.76	1.08	0.32	0.33	0.09	0.00	0.00	0.49
255L3N	12.70	1.05	0.53	0.19	0.07	0.00	0.00	0.08
257L3N	9.16	1.28	0.43	0.14	0.07	0.00	0.00	0.08
258L3N	10.78	1.24	0.27	0.15	0.10	0.00	0.00	0.05
259L3N	10.40	1.56	0.36	0.08	0.08	0.00	0.00	0.20
260L3N	16.05	1.31	0.62	0.18	0.18	0.00	0.00	0.19
261L3N	11.91	1.16	0.41	0.18	0.05	0.00	0.00	0.00

Table 8.18 tabulates the complexity scores of 10 narrative essays written by 10 third year students at the school of Languages.

Table 8.19 the complexity scores of the argumentative essays of 10 third year students at the school of Languages

Students codes	W/T	T/S	DC/T	PP/NP	RC/NP	P phases/NP	IF/NP	NF Clauses, Phrases/T
221L3A	13.51	1.48	0.65	0.05	0.10	0.00	0.02	0.22
222L3A	13.95	1.25	0.38	0.22	0.05	0.01	0.02	0.25
223L3A	13.98	1.18	0.65	0.39	0.16	0.00	0.00	0.43
224L3A	14.86	1.23	0.37	0.15	0.03	0.00	0.02	0.12
225L3A	12.12	1.19	0.33	0.16	0.09	0.00	0.01	0.16
226L3A	13.63	1.46	0.39	0.20	0.03	0.00	0.00	0.17
227L3A	15.41	1.26	0.74	0.18	0.07	0.00	0.00	0.15
228L3A	15.38	1.22	0.64	0.11	0.08	0.00	0.02	0.10
229L3A	11.50	1.03	0.25	0.01	0.00	0.00	0.00	0.09
230L3A	16.57	1.00	0.60	0.18	0.04	0.00	0.00	0.10

Table 8.19 tabulates the complexity scores of 10 argumentative essays written by 10 third year students at the school of Languages.

As with the previous cases, the points of similarity are between the performance of the students in terms of the participial phrases and *to-infinitive* phrases as post-nominal modifiers. The students of the school of languages, like the students of the other schools, have very few instances of these structures with a slight difference between the argumentative and narrative essays. There is also a slight difference between the narrative and argumentative essays in the mean length of T-unit with the argumentative essays having longer T-units. More coordination can be found in the narrative essays but more subordination in the argumentative essays. As far as the phrasal complexity is concerned, despite the fact that the lowest value can be found in the argumentative essays, they tend to be very similar overall in the use of prepositional phrases as post-nominal modifiers. The case is different with the relative clauses, more students have scored slightly higher with the narrative essays. In the case of the non-finite clauses or phrases per T-unit, though the highest score is obtained by the students who have written narrative essays, the argumentative essays still have higher scores if all scores are considered.

Table 8.20 the complexity scores of the narrative essays of 10 fourth year students at the school of Languages

Students codes	W/T	T/S	DC/T	PP/NP	RC/NP	P phases/NP	IF/NP	NF Clauses, Phrases/T
292L4N	11.40	1.82	0.30	0.11	0.09	0.00	0.00	0.23
293L4N	13.14	1.33	0.64	0.20	0.13	0.00	0.00	0.17
294L4N	13.43	1.03	0.43	0.11	0.04	0.00	0.04	0.06
295L4N	14.32	1.33	0.57	0.25	0.14	0.00	0.00	0.11
296L4N	14.71	1.60	0.71	0.16	0.14	0.00	0.05	0.08
297L4N	15.81	1.14	0.59	0.08	0.10	0.00	0.02	0.22
298L4N	9.30	1.18	0.40	0.06	0.11	0.00	0.00	0.10
299L4N	11.46	1.76	0.43	0.23	0.14	0.00	0.00	0.32
300L4N	14.96	1.30	0.65	0.06	0.13	0.00	0.00	0.23
301L4N	18.18	1.29	0.86	0.12	0.12	0.00	0.00	0.36

Table 8.20 tabulates the complexity scores of 10 narrative essays written by 10 fourth year students at the school of Languages.

Table 8.21 the complexity scores of the argumentative essays of 10 fourth year students at the school of Languages

Students codes	W/T	T/S	DC/T	PP/NP	RC/NP	P phases/NP	IF/NP	NF Clauses, Phrases/T
266L4A	14.64	1.47	0.55	0.07	0.04	0.00	0.00	0.00
267L4A	20.65	1.21	0.82	0.15	0.08	0.00	0.04	0.47
268L4A	13.46	1.14	0.75	0.05	0.10	0.00	0.00	0.33
269L4A	18.56	1.17	0.76	0.13	0.00	0.00	0.00	0.50
270L4A	12.92	1.00	0.69	0.18	0.29	0.06	0.00	0.31
272L4A	20.92	1.09	0.67	0.35	0.03	0.00	0.00	0.08
273L4A	12.67	1.33	0.71	0.14	0.05	0.00	0.00	0.04
274L4A	19.90	1.17	0.90	0.17	0.15	0.00	0.00	0.38
275L4A	11.62	1.63	0.23	0.09	0.02	0.00	0.02	0.00
276L4A	16.31	1.78	0.75	0.18	0.10	0.00	0.02	0.25

Table 8.21 tabulates the complexity scores of 10 argumentative essays written by 10 fourth year students at the school of Languages.

As tables 8.20 and 8.21 show, the argumentative essays have longer T-units than the narrative essays and the students' performance in regard to the participial phrases and *to-infinitive* phrases as post-nominal modifiers is very similar with the argumentative essays scoring slightly higher in the case of participial phrases. The students have scored higher on

coordination in the case of narrative essays but those who wrote the argumentative essays have scored higher in the case of subordination. The phrasal complexity in both argumentative and narrative essays in regard to prepositional phrases does not differ very much but in regard to relative clauses, the students who have written narrative essays have, overall, scored higher, though one of the students who wrote an argumentative essay has obtained the highest score (0.29). A number of students who have written argumentative essays have scored high on the non-finite clauses and phrases reaching (0.50).

Table 8.22 is a summary table that shows a summary of the information in the tables above (table 8.10 to table 8.21). The symbol □ means that this structure is higher. The symbol □✗ means that it is a little higher. The symbol S means the structure is similar in the argumentative and narrative essays.

Table 8.22 a summary of the tables (8.10 to 8.21)

Groups	Argumentative								Narrative							
	W/T	T/S	DC/T	PP/NP	RC/NP	P/NP	IF/NP	NF clauses, phrases/NP	W/T	T/S	DC/T	PP/NP	RC/NP	P/NP	IF/NP	NF clauses, phrases/NP
3th year Arts	✓	✓	✓	✓		S	S						✓	S	S	✓
4th year Arts	✓		✓✗	✓		S	S	✓		✓			✓	S	S	
3rd year Basic Education	✓		✓			S	✓	✓		✓		✓✗	✓✗	=		
4th year Basic Education	✓		✓	✓	✓		S	✓		✓				✓✗	S	
3th year Languages	✓		✓	S		S	S	✓		✓		S	✓	S	S	
4th year Languages	✓		✓	S		✓✗	S	✓		✓		S	✓		S	

Looking at table 8.22, one can observe that the argumentative essays seem to be more complex than the narrative essays since on most of the complexity levels like subordination, the use of non-finite clauses and phrases in nominal and adverbial positions, the mean length of T-units, and the use of prepositional phrases as post-nominal modifiers the students who

wrote argumentative essays have scored higher than those who wrote the narrative essays. Two other points need to be highlighted here. The first one is that the coordination index is higher in most of the cases with the narrative essays but, as stated earlier, the excessive use of coordination might only be a means of avoiding more complex language by combining many simple sentences or even fragments with a coordinating conjunction like *and*. The second point is that the relative clauses as post-nominal modifiers also tend to be higher in narrative essays. It is important to note, though, that relative clauses are less complex than prepositional phrases as post-nominal modifiers because they involve only relativization but not reduction. Moreover, what is indicative of the high complexity of argumentative essays in terms of subordination and phrasal complexity is the high length of T-units and low level of coordination. This means that the T-units have tended to be long due to the addition of dependent clauses and prepositional phrases as post-nominal modifiers.

This comparison may have an important pedagogical implication; it is necessary to require students to write argumentative essays in order to help them write more complex language.

8.6.3 the complexity of the 3rd year students' narrative essays compared to the complexity of the 4th year students' narrative essays in each school

Table 8.23 the complexity scores of the narrative essays of 10 third year students at the school of Arts[3]

Students codes	W/T	T/S	DC/T	PP/NP	RC/NP	P phases/NP	IF/NP	NF Clauses, Phrases/T
1A3N	10.30	1.28	0.19	0.05	0.04	0.00	0.00	0.14
2A3N	18.92	1.41	0.92	0.05	0.32	0.00	0.00	0.58
3A3N	10.92	1.20	0.44	0.13	0.07	0.01	0.00	0.25
4A3N	11.63	1.37	0.32	0.05	0.08	0.00	0.02	0.17
5A3N	10.07	1.16	0.41	0.14	0.06	0.00	0.00	0.00
6A3N	12.54	1.05	0.72	0.02	0.18	0.00	0.02	0.39
7A3N	14.16	1.14	0.52	0.13	0.17	0.00	0.02	0.24
8A3N	17.27	1.22	0.50	0.18	0.12	0.00	0.02	0.23
9A3N	15.62	1.71	0.69	0.21	0.14	0.00	0.00	0.10
10A3N	10.45	1.10	0.34	0.01	0.09	0.00	0.00	0.08

This table (8.23) shows the complexity scores of 10 third year students at the school of Arts on narrative essays.

Table 8.24 the complexity scores of the narrative essays of 10 fourth year students at the school of Arts

Students Codes	W/T	T/S	DC/T	PP/NP	RC/NP	P phases/NP	IF/NP	NF Clauses, Phrases/T
58A4N	16.12	1.55	0.41	0.08	0.16	0.00	0.00	0.18
59A4N	14.00	1.32	0.39	0.20	0.05	0.00	0.04	0.24
60A4N	13.30	1.76	0.30	0.28	0.09	0.00	0.00	0.24
61A4N	13.52	1.48	0.42	0.25	0.02	0.00	0.00	0.13
62A4N	19.00	1.28	1.16	0.31	0.18	0.00	0.00	0.34
63A4N	11.05	1.65	0.42	0.19	0.11	0.00	0.00	0.19
64A4N	11.21	1.26	0.38	0.16	0.05	0.00	0.00	0.10
65A4N	17.80	1.59	0.43	0.09	0.07	0.00	0.00	0.40
66A4N	15.79	1.32	0.76	0.21	0.19	0.00	0.00	0.12
67A4N	12.02	1.31	0.41	0.16	0.05	0.00	0.00	0.12

This table 8.24 shows the complexity scores of 10 fourth year students at the school of Arts on narrative essays.

As the tables 8.23 and 8.24 demonstrate, the essays of both third and fourth year students differ in their mean length of T-units. More fourth year students have scored higher than third year students and the length of their T-units ranged between 11.05 W/T to 19.00 W/T while the length of T-units produced by the third year students ranged between 10.07 W/T–18.92 W/T with most of the students within this range scoring less than the fourth year students. The fourth year students have also scored higher on the coordination index as most of them have obtained higher scores than the third year students on the measure T/S.

As for the subordination index or the DC/T, the lowest score in the case of third year student is 0.19 and the highest is 0.92 but in the case of fourth year students the highest score is 1.16 and the lowest is 0.30. However, the two groups' scores in between these figures do not differ very much. Two third year students scored between 0.30 and 0.39 (0.32, 0.34) but three fourth year students scored within this range (0.30, 0.38, 0.39). It is obvious, though, that two of the 0.30s of the fourth year students are closer to 0.4 (0.38 and 0.39) than the 0.30s of the third year students. Yet, looking more closely at the other figures, one can see that more fourth year students scored between 0.4 and 0.49 (five students) but only two third year students scored within this range and more of them (2) have scored between 0.5 and 0.59. When the other ranges are compared to each other 0.6–0.69 and 0.7–0.79, it is obvious that only one third year student falls within the first range (0.69) and one within the second range but only one fourth year student's score (0.76) falls within the second range.

On the phrasal complexity level, despite the fact that the fourth year students have obviously performed better than the third year students in terms of the production of prepositional phrases as post-nominal modifiers, their scores are lower in the case of relative clauses. Both third and fourth year students have scored very low on the participial and *to-infinitive* clauses; yet more students in the third year have scored higher than the fourth year students in the case of *to-infinitive*. The performance of the third year students on the non-finite *to-infinitive* and ing-clauses and phrases seem to be a little better than the performance of the fourth year students though their lowest score is lower than that of the fourth year students (0.00 compared to 0.10)

According to the above comparison (see also table 8.29 below), the fourth year students tend to have written more complex language than the third year students in terms of the mean length of T-unit, subordination, coordination, the use of prepositional phrases as post-nominal modifiers. However, the third year students have produced more non-finite clauses and phrases in nominal and adverbial position (though not to a great extent) and more relative clauses as post-nominal modifiers and these demonstrate, in my opinion, a high level of complexity.

Table 8.25 the complexity scores of the narrative essays of 10 third year students at the school of Basic Education

Students codes	W/T	T/S	DC/T	PP/NP	RC/NP	P phases/NP	IF/NP	NF Clauses, Phrases/T
113B3N	18.11	1.06	0.95	0.29	0.12	0.00	0.03	0.16
114B3N	11.73	1.38	0.36	0.06	0.29	0.00	0.00	0.05
115B3N	11.29	2.50	0.43	0.25	0.04	0.00	0.00	0.03
116B3N	13.45	1.32	0.48	0.12	0.12	0.00	0.00	0.21
117B3N	8.78	1.19	0.22	0.25	0.04	0.00	0.00	0.02
118B3N	11.55	1.07	0.52	0.10	0.05	0.00	0.00	0.17
119B3N	16.13	1.07	0.69	0.11	0.14	0.00	0.03	0.00
120B3N	8.85	1.62	0.32	0.06	0.09	0.00	0.00	0.03
121B3N	13.44	1.30	0.21	0.04	0.10	0.00	0.01	0.19
122B3N	13.73	1.25	0.64	0.26	0.08	0.00	0.00	0.29

This table (8.25) shows the complexity scores of 10 third year students at the school of Basic Education on narrative essays.

Table 8.26 the complexity scores of the narrative essays of 10 fourth year students at the school of Basic Education

Students codes	W/T	T/S	DC/T	PP/NP	RC/NP	P phases/NP	IF/NP	NF Clauses, Phrases/T
139B4N	12.28	1.59	0.51	0.12	0.09	0.00	0.02	0.05
140B4N	15.06	1.07	0.53	0.15	0.08	0.00	0.02	0.34
141B4N	15.03	1.43	0.47	0.14	0.07	0.00	0.00	0.20
142B4N	8.50	1.00	0.38	0.22	0.00	0.00	0.00	0.16
143B4N	15.27	1.43	0.70	0.23	0.14	0.00	0.00	0.13
144B4N	15.68	1.32	0.52	0.27	0.02	0.00	0.00	0.04
145B4N	11.82	1.21	0.35	0.04	0.13	0.08	0.00	0.18
146B4N	12.85	1.11	0.40	0.21	0.05	0.00	0.00	0.00
147B4N	14.91	1.53	0.48	0.07	0.10	0.00	0.00	0.13
148B4N	8.19	1.28	0.05	0.05	0.04	0.00	0.00	0.07

This table (8.26) shows the complexity scores of 10 fourth year students at the school of Basic Education on narrative essays.

Although longer T-units are produced by the third year students (18.11 average words per T-unit), four fourth year students have scored between 15.00 and 15.99 and no third year student has scored in this range, meaning that the fourth year students have scored higher. As for the coordination T/S, the third year students seem to have scored higher on average. If one considers the range between 1.00 and 1.09 respectively, there are three third year students who scored within this range but there are two fourth year students who have scored within this range as well. Also, the scores of one third year and one fourth year student fall within the range between 1.10 and 1.19. The only noticeable difference in performance is that two fourth year students scored between 1.40–1.49 and two between 1.50–1.59 while none of the third year student scored within this range. However, what made the third year students seem to have scored higher is the score 2.50 that was achieved by the student number 115B3N in table 8.25.

The third year students have also scored better in terms of subordination and the production of prepositional phrases as post-nominal modifiers though the difference is not very clear in the case of prepositional phrases. However, the difference is clear in the case of relative clauses: the third year students have done better than the fourth year students. There is also a very slight difference between the two groups in terms of their performance on the participial and *to-infinitive* phrases as post-nominal modifiers; both of the groups have scored very low in this area, though. Both groups' performance on the non-finite clauses and phrases does not seem to be high but the fourth year students have scored slightly higher scores than the third

year students. Another point that attracts attention is that the scores of both groups are homogenous.

Based on the above comparison (see also table 8.29 below), the third year students tend to have used more complex language in terms of subordination, coordination, prepositional phrases as post-nominal modifiers, relative clauses and *to-infinitive* phrases as post-nominal modifiers. The fourth year students have written slightly longer T-units with more non-finite clauses and phrases in nominal and adverbial position. The slightly higher length of T-units of the fourth year students without scoring higher on the complexity measures like prepositional phrases, relative clauses, subordination and coordination might most probably be due to juxtaposing words without having a structured T-unit that has become more complex due to the use of more complexity devices. This might point to the fact that the length of T-unit might not be a good indicator of complexity.

Table 8.27 the complexity scores of the narrative essays of 10 third year students at the school of Languages

Students codes	W/T	T/S	DC/T	PP/NP	RC/NP	P phases/NP	IF/NP	NF Clauses, Phrases/T
250L3N	12.74	2.88	0.30	0.21	0.03	0.00	0.03	0.04
252L3N	12.97	1.30	0.57	0.10	0.15	0.01	0.01	0.14
253L3N	8.73	1.94	0.12	0.13	0.00	0.03	0.03	0.18
254L3N	12.76	1.08	0.32	0.33	0.09	0.00	0.00	0.49
255L3N	12.70	1.05	0.53	0.19	0.07	0.00	0.00	0.08
257L3N	9.16	1.28	0.43	0.14	0.07	0.00	0.00	0.08
258L3N	10.78	1.24	0.27	0.15	0.10	0.00	0.00	0.05
259L3N	10.40	1.56	0.36	0.08	0.08	0.00	0.00	0.20
260L3N	16.05	1.31	0.62	0.18	0.18	0.00	0.00	0.19
261L3N	11.91	1.16	0.41	0.18	0.05	0.00	0.00	0.00

This table (8.27) shows the complexity scores of 10 third year students at the school of Languages on narrative essays.

Table 8.28 the complexity scores of the narrative essays of 10 fourth year students at the school of Languages

Students codes	W/T	T/S	DC/T	PP/NP	RC/NP	P phases/NP	IF/NP	NF Clauses, Phrases/T
292L4N	11.40	1.82	0.30	0.11	0.09	0.00	0.00	0.23
293L4N	13.14	1.33	0.64	0.20	0.13	0.00	0.00	0.17
294L4N	13.43	1.03	0.43	0.11	0.04	0.00	0.04	0.06
295L4N	14.32	1.33	0.57	0.25	0.14	0.00	0.00	0.11
296L4N	14.71	1.60	0.71	0.16	0.14	0.00	0.05	0.08
297L4N	15.81	1.14	0.59	0.08	0.10	0.00	0.02	0.22
298L4N	9.30	1.18	0.40	0.06	0.11	0.00	0.00	0.10
299L4N	11.46	1.76	0.43	0.23	0.14	0.00	0.00	0.32
300L4N	14.96	1.30	0.65	0.06	0.13	0.00	0.00	0.23
301L4N	18.18	1.29	0.86	0.12	0.12	0.00	0.00	0.36

This table (8.28) shows the complexity scores of 10 fourth year students at the school of Languages on narrative essays.

Tables 8.27 and 8.28 demonstrate that the essays written by the fourth year students consist of longer T-units than those written by third year students. What is noticeable here is that the two groups' scores are homogenous i.e. there is not big difference between their marks. Regarding the use of coordination and subordination, although they do not differ very much overall, the third year students have used more coordinated T-units but the fourth year students have used more subordination. This is because (as the tables 8.28 and 8.29 demonstrate) almost all ten fourth year students whose language is analyzed have scored higher than the ten third year students on the production of subordination.

As for the phrasal complexity in terms of the production of prepositional phrases, the third year students seem to have outperformed the fourth year students. However, more fourth year students have produced a higher number of relative clauses than the third year students. Like the previous comparisons, all the fourth and third year students have produced a very low number of participial and *to-infinitive* phrases as post-nominal modifiers. Although one third year student has scored very high on the production of non-finite clauses and phrases (0.49), all the fourth year students have produced more of these phrases and clauses.

The fourth year students have depended on long T-units, subordination, the use of non-finite clauses and phrases, and the production of relative clauses in their noun phrases while the third year students have produced shorter T-units with less subordination. Both groups, like

all the other groups, have produced a very low number of participial and *to-infinitive* phrases as post-nominal modifiers (see table 8.29 below for more clear comparison).

For making this comparison more accessible to the reader, a summary of all the tables (8.23, 8.24, 8.25, 8.26, 8.27, 8.28) is presented in table 8.29. This table compares third year students to fourth year students in three schools (school of Arts, school of Basic Education, school of Languages) in their performance on narrative essays in terms of mean length of T-unit or word per T-unit (W/T), T-units per sentence or the coordination measure (T/S), dependent clauses/ T-unit or the subordination measure (DC/T), prepositional phrase as post-nominal modifiers per noun phrase (PP/NP), participial phrase as post-nominal modifiers per noun phrase (P phrases/NP), *to-infinitive* phrase as post-nominal modifiers per noun phrases (IF/NP) and non-finite clauses and phrases as nominal and adverbial per T-unit (NF Clauses, Phrases/T). The symbol □ points to a higher ratio, □✗ indicates a slightly higher ratio, and S indicates exactly the same ratio.

Table 8.29 a summary of the tables (8.23 to 8.28)

School of Arts									**School of Basic Education**								**School of Languages**							
	W/T	T/S	DC/T	PP/NP	RC/NP	P phases/NP	IF/NP	NF Clauses, Phrases/T	W/T	T/S	DC/T	PP/NP	RC/NP	P phases/NP	IF/NP	NF Clauses, Phrases/T	W/T	T/S	DC/T	PP/NP	RC/NP	P phases/NP	IF/NP	NF Clauses, Phrases/T
3rd year					✓	S	✓✗	✓✗		✓	✓	✓✗	✓		✓✗			✓		✓		S	S	
4th year	✓	✓	✓✗	✓		S			✓					✓✗		✓		✓		✓		S	S	✓

8.6.4 The complexity of the 3rd year students' argumentative essays compared to the complexity of the 4th year students' argumentative essays in each school

Table 8.30 the complexity scores of the argumentative essays of 10 third year students at the school of Arts

Students codes	W/T	T/S	DC/T	PP/NP	RC/NP	P phases/NP	IF/NP	NF Clauses, Phrases/T
33A3A	16.18	1.47	0.68	0.24	0.10	0.00	0.00	0.09
34A3A	12.77	1.05	0.33	0.36	0.01	0.00	0.00	0.03
35A3A	20.05	1.33	0.55	0.10	0.02	0.00	0.00	0.25
36A3A	14.83	1.14	0.64	0.08	0.14	0.00	0.06	0.07
37A3A	20.57	1.40	0.93	0.37	0.16	0.00	0.05	0.00
38A3A	36.55	1.10	2.45	0.14	0.17	0.00	0.00	0.91
39A3A	13.80	1.92	0.40	0.21	0.06	0.00	0.00	0.00
40A3A	9.51	1.18	0.16	0.16	0.02	0.00	0.00	0.09
41A3A	11.86	1.06	0.35	0.19	0.15	0.00	0.04	0.03
42A3A	19.31	1.38	0.69	0.14	0.05	0.00	0.00	0.14

This table (8.30) shows the complexity scores of 10 third year students at the school of Arts on argumentative essays.

Table 8.31 the complexity scores of the argumentative essays of 10 fourth year students at the school of Arts

Students codes	W/T	T/S	DC/T	PP/NP	RC/NP	P phases/NP	IF/NP	NF Clauses, Phrases/T
81A4A	11.64	1.04	0.36	0.30	0.00	0.00	0.01	0.19
84A4A	20.70	1.03	0.47	0.30	0.08	0.00	0.00	0.97
85A4A	11.98	1.53	0.29	0.18	0.05	0.00	0.01	0.12
86A4A	31.67	1.33	1.08	0.19	0.08	0.00	0.00	0.50
87A4A	22.74	1.53	0.91	0.15	0.07	0.00	0.02	0.22
88A4A	18.64	1.25	0.84	0.16	0.13	0.00	0.00	0.20
89A4A	17.33	1.20	1.08	0.18	0.06	0.00	0.00	0.42
90A4A	16.89	1.27	0.68	0.23	0.14	0.00	0.00	0.32
92A4A	12.64	1.57	0.32	0.31	0.11	0.00	0.00	0.05
93A4A	13.16	1.58	0.47	0.12	0.06	0.00	0.00	0.00

This table (8.31) shows the complexity scores of 10 fourth year students at the school of Arts on argumentative essays.

As for the mean length of T-unit, the third year students have scored the highest mark (36.55) compared to that scored by the fourth year students (31.67). However, overall the fourth year students tend have outperformed them. Also, their scores are more homogeneous than the third year students; their scores are closer to each other than the scores of the third year

students starting from 11.64 and ending with only 31.67, but in the case of third year students there is a higher difference between the 9.51 and 36.55. The same applies to the coordination index; the scores of the fourth year students seem to be more homogenous than the scores of the third year students and they seem to have slightly outperformed the third year students. In terms of subordination, the third year students have scored higher.

As regards phrasal complexity, the third and fourth year students are very similar in their performance on the participial and *to-infinitive* phrases as post-nominal modifiers, though the third year students tend to be slightly better than the fourth year students as far as the *to-infinitive* phrases are concerned. The two groups' performance on prepositional phrases and relative clauses does not differ to a great extent but the third year students have scored a little higher on relative clauses, and the opposite is true with regard to the performance on prepositional phrases. The fourth year students have obviously scored higher than the third year students in the performance on the non-finite clauses and phrases.

What is very clear in this comparison (see also table 8.36 below) is that there is a lot of similarity between the third and fourth year students except for the fact that there is more homogeneity in the level of fourth year students in terms of the mean length of T-unit and coordination index, and the language of the fourth year students is more complex in regard to the use of non-finite clauses and phrases.

Table 8.32 the complexity scores of the argumentative essays of 10 third year students at the school of Basic Education

Students codes	W/T	T/S	DC/T	PP/NP	RC/NP	P phrases/NP	IF/NP	NF Clauses, Phrases/T
191B3A	19.46	1.04	0.88	0.09	0.06	0.00	0.04	0.13
192B3A	16.80	1.15	0.53	0.15	0.05	0.00	0.02	0.27
193B3A	13.71	1.23	0.55	0.10	0.06	0.00	0.02	0.21
195B3A	16.35	2.06	0.57	0.15	0.06	0.00	0.01	0.11
196B3A	18.57	1.00	0.57	0.25	0.12	0.00	0.00	0.33
197B3A	25.97	1.20	0.73	0.10	0.10	0.00	0.03	0.00
198B3A	13.81	1.04	0.77	0.11	0.13	0.00	0.04	0.42
199B3A	19.85	1.23	0.74	0.04	0.01	0.00	0.01	0.11
200B3A	23.95	1.11	1.00	0.20	0.13	0.00	0.01	0.30
201B3A	12.13	1.07	0.34	0.10	0.05	0.00	0.00	0.11

This table (8.32) shows the complexity scores of 10 third year students at the school of Basic Education on argumentative essays.

Table 8.33 the complexity scores of the argumentative essays of 10 fourth year students at the school of Basic Education

Students codes	W/T	T/S	DC/T	PP/NP	RC/NP	P phases/NP	IF/NP	NF Clauses, Phrases/T
161B4A	12.64	1.08	0.43	0.14	0.14	0.00	0.00	0.25
162B4A	18.60	1.25	0.80	0.24	0.12	0.00	0.00	0.53
167B4A	11.10	1.35	0.29	0.14	0.07	0.00	0.00	0.13
168B4A	15.22	1.23	0.41	0.13	0.03	0.00	0.02	0.11
169B4A	11.55	1.27	0.47	0.12	0.04	0.00	0.00	0.16
171B4A	20.56	1.13	1.28	0.23	0.16	0.00	0.00	0.56
172B4A	12.64	1.12	0.21	0.30	0.03	0.00	0.00	0.18
173B4A	12.13	1.14	0.29	0.24	0.09	0.00	0.04	0.13
174B4A	14.12	1.14	0.60	0.00	0.17	0.00	0.00	0.28
175B4A	14.54	1.04	0.57	0.11	0.07	0.00	0.02	0.14

This table (8.33) shows the complexity scores of 10 fourth year students at the school of Basic Education on argumentative essays.

The first columns in table 8.32 and table 8.33 (see also table 8.36 below) show that the mean length of the T-units produced by the third year students is much higher than the mean length of the T-units produced by the fourth year students. The third year students seem to have depended on subordination rather than coordination in lengthening their T-units because their coordination index is not higher to a great extent than that of the fourth year students but their subordination index is obviously higher.

Concerning phrasal complexity, the fourth year students have produced more prepositional phrases and slightly more relative clauses, meaning that they have mostly depended on phrasal complexity and the use of non-finite clauses and phrases in adverbial and nominal positions. The two groups did not do well on the use of participial clauses and phrases as post-nominal modifiers but the third year students are better than the fourth year students (though they are also not very good) in the production of *to-infinitive* phrases for post-modification.

Table 8.34 the complexity scores of the argumentative essays of 10 third year students at the school of Languages

Students codes	W/T	T/S	DC/T	PP/NP	RC/NP	P phases/NP	IF/NP	NF Clauses, Phrases/T
221L3A	13.51	1.48	0.65	0.05	0.10	0.00	0.02	0.22
222L3A	13.95	1.25	0.38	0.22	0.05	0.01	0.02	0.25
223L3A	13.98	1.18	0.65	0.39	0.16	0.00	0.00	0.43
224L3A	14.86	1.23	0.37	0.15	0.03	0.00	0.02	0.12
225L3A	12.12	1.19	0.33	0.16	0.09	0.00	0.01	0.16
226L3A	13.63	1.46	0.39	0.20	0.03	0.00	0.00	0.17
227L3A	15.41	1.26	0.74	0.18	0.07	0.00	0.00	0.15
228L3A	15.38	1.22	0.64	0.11	0.08	0.00	0.02	0.10
229L3A	11.50	1.03	0.25	0.01	0.00	0.00	0.00	0.09
230L3A	16.57	1.00	0.60	0.18	0.04	0.00	0.00	0.10

This table (8.34) shows the complexity scores of 10 third year students at the school of Languages on argumentative essays.

Table 8.35 the complexity scores of the argumentative essays of 10 fourth year students at the school of Languages

Students codes	W/T	T/S	DC/T	PP/NP	RC/NP	P phases/NP	IF/NP	NF Clauses, Phrases/T
266L4A	14.64	1.47	0.55	0.07	0.04	0.00	0.00	0.00
267L4A	20.65	1.21	0.82	0.15	0.08	0.00	0.04	0.47
268L4A	13.46	1.14	0.75	0.05	0.10	0.00	0.00	0.33
269L4A	18.56	1.17	0.76	0.13	0.00	0.00	0.00	0.50
270L4A	12.92	1.00	0.69	0.18	0.29	0.06	0.00	0.31
272L4A	20.92	1.09	0.67	0.35	0.03	0.00	0.00	0.08
273L4A	12.67	1.33	0.71	0.14	0.05	0.00	0.00	0.04
274L4A	19.90	1.17	0.90	0.17	0.15	0.00	0.00	0.38
275L4A	11.62	1.63	0.23	0.09	0.02	0.00	0.02	0.00
276L4A	16.31	1.78	0.75	0.18	0.10	0.00	0.02	0.25

This table (8.35) shows the complexity scores of 10 fourth year students at the school of Languages on argumentative essays.

As the tables 8.34 and 8.35 show (see also table 8.36), most of the fourth year students' T-units are longer than those produced by the third year students, and they seem to have depended on subordination rather than coordination in lengthening their T-units because their scores on subordination are higher than the scores of the third year students.

The phrasal complexity of both years does not differ very much in regard to the production of participial and *to-infinitive* phrases as post-nominal modifiers. Nevertheless, there is a difference in the case of prepositional phrases and relative clauses with the third year students scoring higher than the fourth year students in terms of prepositional phrases but the fourth year students scoring higher than the third year students in terms of relative clauses. As for the use of non-finite clauses and phrases, the fourth year students seem to have done better than the third year students.

For making this comparison more accessible to the reader, a summary of all the tables (8.30, 8.31, 8.32, 8.33, 8.34, 8.35) is presented at table 8.36. This table compares third year students to fourth year students in three schools (school of Arts, school of Basic Education, school of Languages) in their performance on argumentative essays (with regard to which measures the comparison includes see the description of table 8.29 above).

Table 8.36 a summary of the tables 8.30 to 8.35

School of Arts								
	W/T	T/S	DC/T	PP/NP	RC/NP	P phases/NP	IF/NP	NF Clauses, Phrases/T
3rd year			✓		✓✗	S	✓✗	
4th year	✓	✓		✓✗		S		✓

School of Basic Education								
	W/T	T/S	DC/T	PP/NP	RC/NP	P phases/NP	IF/NP	NF Clauses, Phrases/T
3rd year	✓	✓	✓			S	✓✗	
4th year				✓	✓✗	S		✓

School of Languages								
	W/T	T/S	DC/T	PP/NP	RC/NP	P phases/NP	IF/NP	NF Clauses, Phrases/T
3rd year				✓			S	
4th year	✓	✓	✓		✓	✓✗	S	✓

8.7 Conclusion

This chapter has dealt with second language or foreign language complexity and complexity measures that are in many cases used as metrics for language proficiency and language development. It has defined complexity and shown that its classification is complicated, having a hierarchical shape. It consists of two main types, cognitive complexity (difficulty) and absolute complexity (complexity). The focus in this chapter is on absolute complexity

and, more specifically, on its syntactic structure complexity as a subtype of linguistic complexity. If we go further with the classification, the core focus is on the sentential, clausal and phrasal complexity of a number of essays written by the students of three schools at two universities in Iraqi Kurdistan.

This chapter answered a number of questions. The first question, about the measures of complexity in previous research, is answered by showing that there are a number of measures used in measuring complexity in L2 or FL writing. These measures included (1) measures of overall complexity such as the mean length of production unit (e.g. sentences, clauses, T-units, C-units etc.), (2) measures of coordination such as coordinated clauses/clauses, T-units/sentence, etc. (3) measures of subordination that included clauses/other production units (e.g. T-units, C-unit, AS-unit), dependent or subordinate clauses/T-units, etc. (4) measures of phrasal and clausal complexity such as dependents/(noun, verb) phrase, complex nominals/T-units, syntactic arguments/clause, etc. A number of these measures, for example, mean length of T-unit, T-units/ sentences, dependent clauses/ T-units have been successfully applied to the data used in the present chapter as an answer to the second question posed here. However, a number of new measures, which involve measuring post-modification, have been used to operationalize the phrasal complexity. In addition to this, a number of non-finite phrases and clauses divided by the total number of T-units is used in an attempt to measure how many non-finite clauses and phrases are there in a T-unit.

As for question 3, a number of students have made a balance between accuracy and complexity. Some others could not keep this balance and either did well on accuracy or complexity. All students' performance on some phrasal complexity measures like the *to-infinitive* and participial phrases is very low. One could very rarely find examples of these structures in the data.

The argumentative essays are more complex than the narrative essays in most of the cases with remarkably longer T-units. Their complexity is not only on the sentential and clausal level but the phrasal level as well. Contrary to this, Foster and Skehan (1996) found that narrative essays were the least accurate and the most complex compared to their other two tasks (Decision-making and Personal-information exchange).

The ratio of T-units to sentences i.e. the ratio of coordination, however, is higher in the narrative essays in more cases. This could very much be due to the fact that the narrative essays are a conversational type of writing where simple noun phrases and simple sentences

that are connected only through coordination are the common features. The amount of coordination may only be a means of avoiding complex language that is mostly represented in phrasal complexity and the use of non-finite clauses.

The third year students' complexity scores compared to the fourth year students' complexity scores (in all three schools: school of Arts, school of Basic Education, and school of Languages) on the narrative essays seem to differ slightly on some measures but considerably on some others. For example, the fourth year students at the school of Arts have produced longer T-units than the third year students with slightly less subordination but obviously more coordination. Also, a not significant difference can be found between the third and fourth year students at the school of Basic Education in terms of their production of prepositional phrases as post-nominal modifiers, but the third year students have scored noticeably higher than the fourth year students as far as their production on relative clauses is concerned. Another example can be the fourth year students at the school of Languages who have scored higher on the mean length of T-units, subordination, relative clauses and non-finite clauses and phrases in adverbial and nominal positions but whose performance does not differ very much from the performance of the third year students in the case of participial phrases and clauses.

The students' performance on argumentative essays also yielded mixed results. For example, the third year students at the school of Arts slightly outperformed the fourth year students on the *to-infinitive* phrases and relative clauses as post-head modifiers in noun phrases but the fourth year students' language tends to be more complex in terms of non-finite clauses and phrases in nominal and adverbial positions. As for the school of Basic Education, the third year students produced longer T-units, more subordination, and a few more *to-infinitive* as post-head modifiers. However, the fourth year students produced more non-finite clauses and phrases in the nominal and adverbial position. The production of the fourth year students at the school of Languages is more complex than the production of the third year students in all cases except for the case of prepositional phrases where the third year students have scored higher. It is important to note here that I have usually taken complexity (but not in all cases, like the case with excessive sentential coordination) as an indicator of quality, so in general, the more complexity there is in the essay, the more advanced the student.

Taking all the discussion above into consideration, one can conclude three important points. The first one is that some kinds of complexity indicate a better grasp of the language, like phrasal complexity and the use of non-finite clauses and phrases but others do not, like the excessive use of coordination. The second one is that complexity is more appropriate to some

genres than others as is the case with argumentative writing when compared to the narrative. The third point is that in any case, one has to offset complexity against accuracy in judging how good or bad an essay is.

Another highly important remark also worth making here which is that complexity can better be operationalized by measuring the occurrence of its different devices and structures. It cannot only be measured by a general measure but specific measures because not all subordination conjunctions and not all subordinating structures have the same linguistic and cognitive complexity. This is the reason why different measures like the measure of non-finite clauses in nominal and adverbial positions and the measure of relative clauses per noun phrase are used in this research besides the general measure of subordination. It is not only syntactic complexity that is important but lexical complexity as well and this the topic of the next chapter.

Notes

1. Bultè and Housen (2012:31) with these measures also list measures of morphological and lexical complexity.
2. The Syntactic Complexity Analyzer is computer software that can analyze text automatically in terms of syntactic complexity using 14 measures. It is designed by Xiaofei Lu (see Lu, 2012 and Lu, 2010).
3. These figures are the same as above but they have been repeated in this section for easy access and reading.
4. These cases existed but they were very few in number.

Chapter Nine

Lexical Complexity[1]

9.1 Introduction

The lexicon can be regarded as an important component in all language learning tasks in general and both spoken and written language text production tasks like essays and oral narratives etc. in particular. This significance is not only from the perspective of language teachers and researchers but learners as well. This is pointed out by Leki and Carson (1994 cited in Engber, 1995) who carried out a survey of 128 non-native speakers of English (undergraduates) enrolled in an English for academic purposes course. The students ranked vocabulary as the first thing they liked to learn in the language component of the course. Lexicon is also considered as "a significant component in both the construction and interpretation of meaningful text" (Grabe, 1985 cited in Egnber, 1995:141). Vocabulary has also been the focus of research in terms of assessment of lexical complexity (e.g. Kyle and Crossley 2015; Azodi *et al*., 2014; Šišková, 2012; Lu, 2012; V. Johansson, 2008; Wolfe-Quintero *et al.,* 1998; Egnber, 1995; Laufer and Nation, 1995; Linnarud, 1986). Lexical complexity is described in the literature as having three dimensions, namely lexical density, lexical diversity, and lexical sophistication (Lu, 2012; Read, 2000; see also V. Johansson, 2008 and Lindqvist *et al*. 2013 for a discussion on the topic). Read (ibid) has referred to another dimension, which is the proportion of errors (Read, 2000; Lindqvist *et al*. 2013).

It is clear from the literature that measuring these dimensions is not easy because of lack of tools of automated analysis. This has been reflected in the small sets of data used or the use of a small number of measures. In this chapter, I consider a recently developed computer tool (the Lexical Complexity Analyzer or LCA, see Lu, 2012) that automatically measures these three dimensions making use of 25 measures that have been reported in the literature. Thus, this chapter reviews these dimensions and measures them in the data of the present study using this tool. However, like syntactic complexity the notion of lexical complexity is also multilayered and difficult to measure manually. It is even difficult to measure it automatically by computer software in the case of learner corpus the present study, which is full of spelling and morphological errors. Nevertheless, rather than ignore this feature completely, I have chosen to take a program (like the Lexical Complexity Analyzer) that claims to measure lexical complexity.

Taking the above view into consideration, this chapter only explore the notion of lexical complexity in learner corpus and hence it is merely an exploratory chapter. This chapter can notify other researchers of the limitations and advantages of the LCA. It could give the reader a good background about lexical complexity and its possible measures and the weak points and strong points of measuring lexical complexity automatically in a highly deviant learner corpus.

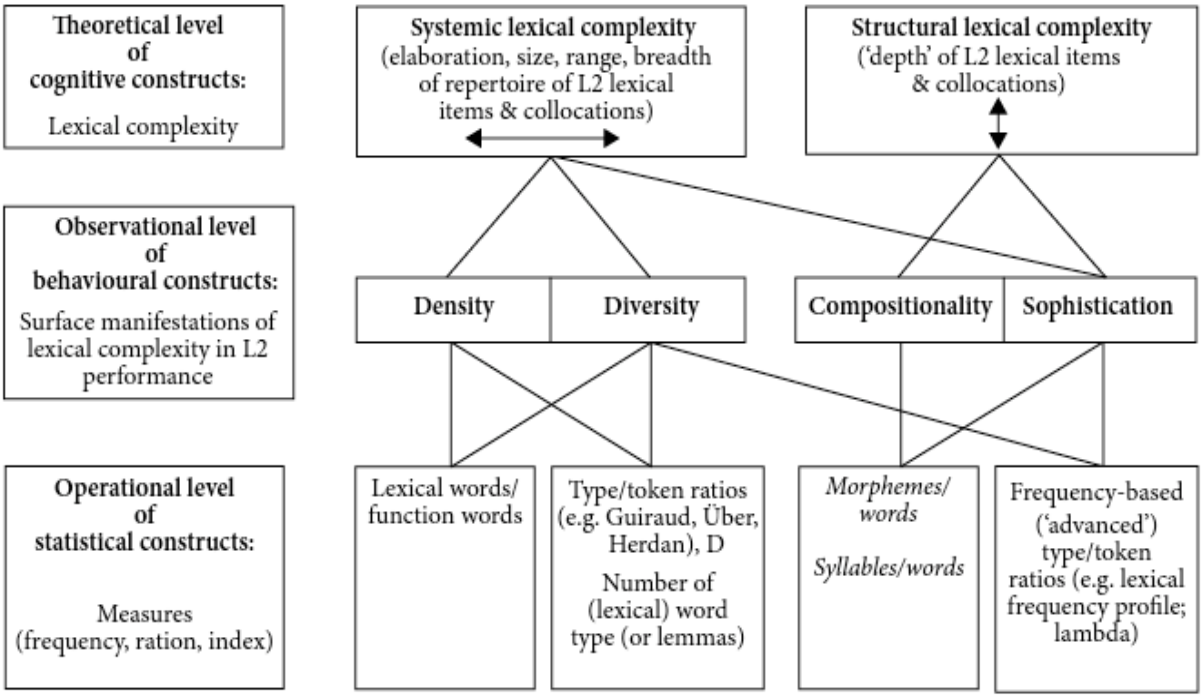
This chapter first provides an account of the three dimensions of lexical complexity in section 9.2. A description of the software (with its all measures) and a calculation of its reliability are also provided in section 9.3. Section 9.4 demonstrates the results of applying this automated analysis to the data of the present study and makes a comparison of the results. The chapter ends with a conclusion.

It is intended to answer the following questions through this chapter: (1) Can the newly developed software LCA (Lu, 2012) be used in the analysis of the lexical complexity of learner data? (2) How does lexical complexity differ from year three to year four (in the case of the three schools, school of Arts, Basic Education, and Languages)? (3) How does lexical complexity differ from argumentative to narrative essays? The answers to these questions, I believe, are important and will most probably contribute to the understanding of lexical complexity in learner corpora, and the way lexical complexity can be automatically measured.

9.2 Lexical complexity and its measures

As mentioned above, lexical complexity is a construct with three dimensions. Bulté and Housen, (2012:28) have demonstrated these dimensions (together with an additional one i.e. compositionality, which is "the number of formal and semantic components of lexical items (e.g. phonemes, morphemes, denotations)" (ibid) and their relation to each other on different layers in the following figure.

Figure 9.1 Bulté and Housen's, (2012:28) classification of lexical complexity on different layers.



As the figure shows, the three dimensions are interconnected and can be operationalized by different measures. The following sections are devoted to the explanation of these dimensions together with a number of their measures.

In addition to all these dimensions, Linnarud (1986) has introduced another dimension i.e. lexical individuality or originality. This dimension can be calculated by dividing the total number of lexical words that are only used by a given writer in a given group by the total number of lexical words produced by this writer (see also Laufer and Nation, 1995 and Wolfe-Quintero *et al.* 1998 for a discussion on this measure). Based on the ideas in the literature and mostly based on the figure above, one can define lexical complexity as the degree to which a given writer uses a *good* number of *varied*, *sophisticated* lexical words and words that are unique to them to elaborate the text as required in order to communicate the meaning that is intended to be conveyed by the text, and it can be operationalized by a number of measures.

9.2.1 Lexical density

Lexical density is the most traditional of the measures that I am using. It is "the ratio of the number of lexical (as opposed to grammatical) words to the total number of words in a text"

(Lu, 2012:191). The term was first coined by Ure (1971 as cited in Lu, 2012 and Read, 2000). The lexical density of the text will increase with the increase of lexical words like verbs, nouns, adjectives and adverbs. As Ure (1971 cited in Lu, 2012 and Read, 2000) has indicated, written texts have a higher lexical density than spoken texts. This is because of the packaging of information in noun phrases in written language and the use of multiple clauses in spoken language (Halliday and Martin, 1993). V. Johansson (2008:65) calls conveying more information through content words as "information packaging".

Although lexical density is quite a simple notion, it does require a decision to be taken about what constitutes open and closed class items. As highlighted in Lu (2012), the types of lexical adverbs counted by O'Loughlin (1995) are adverbs of time, manner, and place but those counted by Engber (1995) are the adverbs derived from adjectives particularly the ones with the suffix –ly. Another controversial example is provided by V. Johansson (2008), which is the use of a phrasal verb like *turn up.* She pointed out that Halliday (1985) has counted it as a unit (one lexical item) but Ure (1971) has considered it as consisting of one lexical word (turn) and one grammatical word (up). An interesting point made by Halliday (1985 as cited in V. Johansson) is that the notion of lexicality is a continuum where some items are clearly content words, some others are clearly grammatical but some are in between like prepositions and modal adverbs (e.g. always, perhaps).

When measuring lexical density, it is very important to define exactly what content words are and what grammatical words are. Because I am using the LCA, which has also been used by Lu (2012:192), I have to quote his definition of content words: "as nouns, adjectives, verbs (excluding modal verbs, auxiliary verbs, 'be,' and 'have'), and adverbs with an adjectival base, including those that can function as both an adjective and adverb (e.g., 'fast') and those formed by attaching the –ly suffix to an adjectival root (e.g., 'particularly')". The formula for lexical density can be written as follows:

Lexical density = Total number of lexical words / total number of all words in the text

9.2.2 Lexical sophistication

Milton (2009:131) defines measurement of lexical sophistication as "a calculation of the proportion of infrequent words in a text" (see also V. Johansson, 2008). Whether a word is frequent or infrequent is a relative matter and depends on the way the researcher defines frequency. For example, Linnarud (1986) defines infrequent words as those that have been

introduced at grade nine and the grades after it in the system of education in Sweden. However, for Lu (2012), as will be discussed in the section on LCA, words are infrequent if they are not within the list of the 2000 most frequent words in the British National Corpus (BNC). Laufer and Nations (1995) have referred to this problem in the measures of lexical sophistication in detail and argued that it is very important for the researcher to define infrequent words according to the level of the students. In this case, automatic tools that have been devised for one context might not be a good means of measuring lexical sophistication in another. However, this leads to the necessity for manual calculations of advanced words in the texts under investigation, and this is quite labor-intensive and time-consuming.

According to these discussions, the researcher has to find a balance between the advantages and disadvantages of automatic and manual research of lexical sophistication. A choice has to be made between either a list of words according to the level of students and a small set of data or 'another-context-based' list of sophisticated words and a large set of data. In this research study, I preferred to use the automatic analysis of lexical sophistication with the LCA, although the sophistication of words is defined in terms of the British National Corpus (BNC). I believe that since all the data are being compared with the same list of sophisticated words, it is a kind of standardization of comparison which impacts all the students in the same way. Other researchers have also used automatic tools for analyzing lexical sophistication, such as Kyle and Crossley (2015), who have used the Tool for the Automatic Analysis of Lexical Sophistication (TAALES) which is available online. They have also referred to different researchers who have used tools such as VocabProfile, Linguistic Inquiry and Word Count (LIWC) and Coh-Metrix for studying L2 lexical acquisition. In addition, Lindqvist *et al.* (2013) have used a tool for operationalizing the lexical sophistication of the language of Swedish learners learning French and Italian.

Concerning the measures of lexical sophistication, based on the above definition, Wolfe-Quintero *et al.* (1998) list and explain the following measures:

1. Lexical sophistication-1 (SLW/LW) = the total number of sophisticated lexical words divided by the total number of lexical words (a token/token ratio).
2. Lexical sophistication-2 (SWT/WT) = the total number of sophisticated word types divided by the total number of word types (a type/type ratio).
3. Verb sophistication (SVT/V) = total number of sophisticated verb types divided by the total number of verbs. (a type/token ratio)

It is worth mentioning that researchers use various statistical means to compensate for things like differences in text length. For example, for the third measure, Wolfe-Quintero *et al.* (ibid) recommend the use of the measure developed by Chaudron and Parker (1990), which is SVT^2/V or $SVT/\sqrt{2V}$, and which is adapted from the measure developed by Carrol (1967).

Lu (2012) has used these measures and their variants resulting in the use of five measures. These will be listed with the studies that have used them in the section on the LCA.

9.2.3 Lexical diversity

Lexical diversity (also called 'lexical variation', see Lu, 2012) can be defined as "a measure of the number of different words used in a text" (Djigunović and Krevelj, 2011:253). Hence, the lexical diversity of a piece of written or spoken language can simply be operationalized by counting the number of different words (NDW) in it; the more types there are in a text, the more diverse it is. This is very well indicated in Lu (2012). However, for matters of comparison between texts of different lengths, this measure does not work well because it does not take into account the length of the sample the lexical diversity of which is being measured. Lu (ibid) points to this issue and proposes a way of treating it. Researchers like Thordardottir and Ellis Weismer, (2001 as cited in Lu, 2012:193) have "truncated" their samples to make them of the same length, equal to the length of the shortest sample. This might not be a good idea as this "truncation" might skew the results and result in the waste of data, as argued by Malvern *et al*. 2004 (cited in Lu 2012).

As an alternative to this, a type/token ratio (TTR) measure is introduced (by Templin, 1957 as pointed out by Šišková, 2012 and Lu, 2012, see also V. Johansson, 2008), which is the number of different words (types) to the number of all words (tokens) in the sample (see, Šišková, 2012; Lu, 2012; Iwashita, 2010; Wolfe-Quintero *et al.* 1998). Despite the fact that this measure is widely used, it has been criticized again for not taking the matter of sample length into account (Šišková, 2012; Lu, 2012; Iwashita, 2010; Wolfe-Quintero *et al.* 1998). The problem with this measure is that when the text gets longer, the ratio of types to tokens decreases (Lu, 2012). Or, as Wolfe-Quintero *et al.* (1998: 106) explain "this type/token measure doesn't discriminate between a writer who uses a few types in a short composition and a writer who uses more types in a longer composition". Tweedie and Baayen (1998) and McCarthy (2005) as cited in Šišková (2012) have indicated different ways of dealing with this by (1) having samples with a similar length through time or word limit, or (2) truncating samples to make them only as long as the shortest sample or (3) dividing the sample into

subsamples of a set length and computing the average TTR of all subsamples. Lu (2012) refers to this average TTR as mean segmental TTR (MSTTR). Nevertheless, these procedures affect the validity of the analysis (Šišková, 2012).

To minimize the effect of text length, a number of new transformations of this ratio have been put forward. In Wolfe-Quintero *et al.* (1998) a number of these transformed measures have been described with examples of how they better discriminate between levels of word diversity. The measures are WT/√2W= the number of different word types divided by the square root of two times the total number of words, which is called 'corrected TTR' (see Lu, 2012) and WT^2 /W= word types squared divided by the total words (developed by Carroll 1967 and Chaudron and Parker, 1990 respectively as cited in Wolfe-Quintero *et al.* 1998). Because I am also choosing those measures in the LCA that take text length into account, I will here present Wolfe-Quintero *et al*'s. (1998) table of hypothetical results to compare the validity of the three measures (the original TTR and its two transformations):

Table 9.1 the measure of TTR and its variants for a number of hypothetical writers

A hypothetical text	Word types	Words	WT/W or (TTR)	WT/√2W	WT^2 /W
A	20	100	0.05	1.41	4.00
B	25	100	0.25	1.77	6.25
C	10	40	0.25	1.12	2.50
D	20	80	0.25	1.58	5.00
E	10	20	0.50	1.58	5.00
F	20	40	0.50	2.24	10.00

Taken from Wolfe-Quintero *et al.,* (1998:107)

This table shows the calculation of three measures (a TTR and its variants) for a number of hypothetical essays. It is only intended for comparing the measures with each other. As argued by Wolfe-Quintero *et al.* (1998), this table shows how the length of the sample does not have any role when comparing the type/token ratio having the same proportion. They cite texts B, C and D as examples of this, as they have the same TTR results although their lengths are very different. The hypothetical text has not been rewarded for having more words. Also, notice, as criticized by Wolfe-Quintero *et al.* (1998), that the lower the length of a sample, the higher the ratio of TTR when the types remain the same, meaning that the

hypothetical writer has been rewarded instead of being penalized. This becomes clear by comparing A, D and F. However, according to the results of the two other measures, the hypothetical samples B and F have the highest lexical variation, higher than the other samples because B has the highest number of types and tokens and F has a fewer types but out of a small number of tokens.

Although these transformations have been made to the TTR to normalize the effect of sample size, there is still a need for a measure that can better fit the purpose. Accordingly, a new measure called the D measure has been developed which is dependent on "find[ing] the best fitting curve to model the TTR in the text" (Šišková, 2012:29). This is because, as quoted in Lindqvist *et al*. (2013:110), it computes "TTRs for samples of different text lengths, ranging from samples of 35 words to samples of 50 words, which are taken randomly from the text". Even after this, another new model has been developed that is based on "random sampling" and software has been designed for these calculations called the vocd (Šišková, 2012:29). These two models are different and they are referred to in the literature as two different measures (ibid). Šišková, (2012:29) shows the calculations of these two measures in the following table and the names of the scholars who developed them.

Table 9.2 The D and Vocd-D measures

Name (Author)	*Year*	*Formula*	*Notes*
D (Malvern & Richards)	1997	$TTR = (2/DN)\,[(1 + DN)^{1/2} - 1]$	The final value of LD is determined by adjusting D until the equation converges on the value of the TTR.
Vocd-D (McKee et al.)	2000	Calculated with the use of dedicated vocd software	Blends curve fitting and sampling Final values tend to range between 10 and 100, with higher values indicating greater diversity.

It is, however, worth mentioning that some studies have not found any statistical relation even between the D measure and the level of writing proficiency (e.g. X. Wang, 2014). Also, although V. Johansson (2008) has used the vocd in her comparative study of lexical density and lexical diversity in the Swedish learners' narrative and expository texts, she has referred to the criticism made by Daller *et al*. 2003 against it, who preferred to use the square root of TTR. Moreover, Lindqvist *et al* (2013) also highlighted the criticism made by McCarthy and Jarvis (2007) who concluded that "it still retains a certain degree of sensitivity to text length"

(Lindqvist *et al*, 2013:110) and that it is the text length that decides which measures are useful.

The TTR was then used for specific word classes; for example, all lexical words, verbs, nouns, adjectives and adverbs. Casanave (1994), for example, has used the ratio of the total number of different lexical word types to the total number of lexical words. McClure (1991 as cited in Engber, 1995) has calculated the noun, verb, adjective and adverb diversity by extending the formulae of TTR to these parts of speech.

As the above discussion and description of tools show, there is a range of tools that measure different things and use different statistics. Some of these tools have been around for a long time and some are quite new. These tools have all been brought together by Xiaofei Lu (see Lu, 2012) in readily available software called the Lexical Complexity Analyser (LCA). As far as I know, this tool, which has only been around for six years, has not been extensively used as yet and thus it is useful to apply it to my data.

9.3 Lexical complexity analyzer (LCA)

LCA is a computer tool that can be downloaded free of charge from http://www.personal.psu.edu/xxl13/downloads/lca.html (Lu, 2012). It has been developed by Xiaofei Lu. This tool works in both single mode and batch mode, and it has the potential of application to large sets of data. In the single mode, one can only upload up to two texts, which one can compare. The files have first to be 'part of speech' tagged and lemmatized. However, the batch mode is very useful and easy to use. The researcher can upload up to 100 .txt files at a time for lexical complexity analysis with not less than 50 words and not more than 10000 words in each file [2,3]. The program analyzes the texts and provides results in the form of a comma separated value (CSV) files that can be downloaded and imported into spreadsheets or statistical packages. The tool calculates lexical density, diversity and sophistication applying 25 measures of those used by researchers in their calculation of lexical complexity. Lu (2012) uses the definition stated in section 9.2.1 above for lexical words. As for his definition of word frequency, he states that words are considered infrequent if they are not in the list of the most frequent 2000 words from the British National Corpus (BNC) and the different forms of the same word are considered only as one type. The following are the measures together with their calculations and the authors who have used them in the literature:

Table 9.3 the measures of lexical density and lexical sophistications in the LCA as cited in Lu (2012:193)

Measure	Code	Formula	Examples
Lexical Density	LD	N_{lex}/N	Engber (1995)
Lexical Sophistication-I	LS1	N_{slex}/N_{lex}	Linnarud (1986), Hyltenstam (1988)
Lexical Sophistication-II	LS2	T_s/T	Laufer (1994)
Verb Sophistication-I	VS1	T_{sverb}/N_{verb}	Harley & King (1989)
Corrected VS1	CVS1	$T_{sverb}/\sqrt{2N_{verb}}$	Wolfe-Quintero et al. (1998)
Verb Sophistication-II	VS2	T^2_{sverb}/N_{verb}	Chaudron & Parker (1990)

Table 9.4 the measures of lexical variation in the LCA as cited in Lu (2012:195)

Measure	Code	Formula
Number of Different Words	NDW	T
NDW (first 50 words)	NDW–50	T in the first 50 words of sample
NDW (expected random 50)	NDW–ER50	Mean T of 10 random 50-word samples
NDW (expected sequence 50)	NDW–ES50	Mean T of 10 random 50-word sequences
Type–Token Ratio	TTR	T/N
Mean Segmental TTR (50)	MSTTR–50	Mean TTR of all 50-word segments
Corrected TTR	CTTR	$T/\sqrt{2N}$
Root TTR	RTTR	$T/\sqrt{N}$
Bilogarithmic TTR	LogTTR	$LogT/LogN$
Uber Index	Uber	$Log^2 N/Log(N/T)$
D Measure	D	Based on D in Equation (1)
Lexical Word Variation	LV	T_{lex}/N_{lex}
Verb Variation-I	VV1	T_{verb}/N_{verb}
Squared VV1	SVV1	T^2_{verb}/N_{verb}
Corrected VV1	CVV1	$T_{verb}/\sqrt{2N_{verb}}$
Verb Variation-II	VV2	T_{verb}/N_{lex}
Noun Variation	NV	T_{noun}/N_{lex}
Adjective Variation	AdjV	T_{adj}/N_{lex}
Adverb Variation	AdvV	T_{adv}/N_{lex}
Modifier Variation	ModV	$(T_{adj} + T_{adv})/N_{lex}$

The only measure that is not calculated with this tool is D (Haiyang Ai, Personal Communication), which is mentioned here because it has been used in the study where these measures are cited.

As this is a newly invented tool, questions obviously arise as to how reliable it is. To test this, I have compared calculations of types and tokens with that made by AntConc 3.4.3m as AntConc is a widely recognized source of information. I discovered that there was a correlation of 0.996 for types and 0.998 for tokens.

9.4 Data analysis and discussion

This section describes the process of applying the LCA to my corpus of students' essays. A sample of the data (280 texts) has been analyzed using the LCA and all the measures in tables 9.3 and 9.4 have been applied as one package except the measure D which, as mentioned earlier, is not within the package of measures of this software. The following subsections show the results and compare them to each other, in addition to giving a description of the sample.

9.4.1 The sample

Because the original subsamples are not equal, I decided to take an equal number of students from each year from each school for this exercise. The following table shows the number of students used.

Table 9.5 the sample of the data analyzed for this chapter

School of Arts/ University of Dohuk				School of Basic Education/University of Dohuk				School of Languages/ University of Zakho			
3rd year		4th year		3rd year		4th year		3rd year		4th year	
N	A	N	A	N	A	N	A	N	A	N	A
25	25	25	25	25	25	25	25	20	20	20	20

A= argumentative N= Narrative

9.4.2 The comparison of the narrative essays and argumentative essays of each year of each school

Tables 9.6 to 9.17 show the results of applying the LCA to the sample of the data that is summarized in table 9.5. Moreover, for making the comparisons between tables easier and clearer, the averages in all these tables have been put in one table (9.18).

Table 9.6 the results of the lexical density, diversity and sophistication of the narrative essays of 3rd year students, school of Arts, University of Dohuk

FILENAME	LD	LS1	LS2	CVS1	MSTTR	CTTR	LV	CVV1	NV	ADJV	ADVV	MODV
1Arts3N.txt	0.44	0.2	0.16	0.61	0.71	5.73	0.65	0.64	0.56	0.12	0.07	0.19
2Arts3N.txt	0.45	0.05	0.08	0.17	0.73	5.49	0.45	0.56	0.53	0.12	0.09	0.21
3Arts3N.txt	0.37	0.15	0.17	0.26	0.74	6.26	0.64	0.67	0.7	0.11	0.05	0.17
4Arts3N.txt	0.47	0.28	0.2	0.44	0.73	6.01	0.66	0.59	0.51	0.11	0.06	0.17
5Arts3N.txt	0.53	0.1	0.2	0.22	0.67	3.98	0.33	0.31	0.26	0.06	0.05	0.11
6Arts3N.txt	0.41	0.17	0.18	0.4	0.7	5.35	0.56	0.52	0.5	0.1	0.09	0.19
7Arts3N.txt	0.44	0.08	0.11	0.1	0.67	4.97	0.59	0.54	0.47	0.1	0.03	0.12
8Arts3N.txt	0.5	0.16	0.22	0.33	0.73	5.39	0.69	0.58	0.48	0.12	0.05	0.17
9Arts3N.txt	0.46	0.07	0.1	0.12	0.65	4.15	0.49	0.39	0.3	0.12	0.05	0.18
10Arts3N.txt	0.45	0.16	0.17	0.29	0.72	5.46	0.44	0.55	0.6	0.12	0.1	0.22
11Arts3N.txt	0.46	0.23	0.26	0.36	0.73	5.23	0.54	0.47	0.46	0.12	0.04	0.16
12Arts3N.txt	0.48	0.17	0.19	0.36	0.76	5.85	0.54	0.55	0.58	0.1	0.07	0.16
13Arts3N.txt	0.48	0.27	0.27	0.44	0.74	7.03	0.67	0.68	0.7	0.09	0.08	0.17
14Arts3N.txt	0.41	0.18	0.2	0.64	0.67	5.42	0.52	0.56	0.56	0.14	0.06	0.2
15Arts3N.txt	0.46	0.14	0.17	0.33	0.7	4.4	0.6	0.5	0.41	0.1	0.06	0.16
16Arts3N.txt	0.45	0.14	0.18	0.63	0.72	5.19	0.57	0.52	0.45	0.1	0.06	0.15
17Arts3N.txt	0.47	0.12	0.13	0.58	0.75	5.08	0.68	0.57	0.52	0.14	0.07	0.21
18Arts3N.txt	0.41	0.19	0.22	0.78	0.65	4.49	0.51	0.45	0.44	0.08	0.04	0.13
19Arts3N.txt	0.43	0.17	0.23	0.63	0.77	5.82	0.65	0.57	0.58	0.11	0.05	0.16
20Arts3N.txt	0.45	0.17	0.18	0.52	0.77	6.51	0.64	0.62	0.64	0.11	0.1	0.21
21Arts3N.txt	0.41	0.25	0.23	0.47	0.7	4.92	0.47	0.57	0.66	0.08	0.05	0.12
22Arts3N.txt	0.48	0.23	0.24	0.58	0.72	5.13	0.4	0.4	0.41	0.07	0.06	0.12
23Arts3N.txt	0.45	0.18	0.21	0.71	0.69	5	0.51	0.49	0.52	0.1	0.07	0.17
24Arts3N.txt	0.42	0.25	0.26	0.66	0.66	4.32	0.5	0.44	0.41	0.09	0.04	0.13
25Arts3N.txt	0.47	0.13	0.2	0.36	0.75	5.33	0.64	0.47	0.44	0.1	0.06	0.16
Average	**0.45**	**0.17**	**0.19**	**0.44**	**0.71**	**5.30**	**0.56**	**0.53**	**0.51**	**0.10**	**0.06**	**0.17**

Table 9.6 shows the calculations of 12 measures that have been applied to the sample of 25 narrative essays written by the third year students/ school of Arts, University of Dohuk. The measures are Lexical Density (LD), Lexical Sophistication-1 (LS1), Lexical Sophistication-2 (LS2), Corrected Verb Sophistication-1 (CVS1), Mean Segmental Type/Token Ratio (MSTTR), Corrected Type/Token Ratio (CTTR), Lexical Word Variation (LV), Corrected Verb Variation-1(CVV1), Noun variation (NV), Adjective Variation (ADJV), Adverb variation (ADVV), and Modifier variation (MODV).

Only these measures have been chosen from the sum of 25 measures included in the LCA for reason of limited space. The choice of these measures is based on two points: (1) variation i.e. the measures chosen are varied to include the three aspects of lexical complexity (lexical density, lexical diversity, and lexical sophistication) and also to include all lexical word classes: verbs, nouns, adjectives, and adverbs (2) measure sensitivity to sample length. (Notice that the corrected versions of measures like CTTR, CVV1, CVS1 and MSTTR are used since they are modified versions of their original measures and they take sample length into account. This is because the samples are not of equal length.)

Table 9.7 the results of the lexical density, diversity and sophistication of the argumentative essays of 3rd year students, school of Arts, University of Dohuk

FILENAME	LD	LS1	LS2	CVS1	MSTTR	CTTR	LV	CVV1	NV	ADJV	ADVV	MODV
33Arts3A.txt	0.48	0.16	0.17	0.23	0.65	4.19	0.58	0.43	0.37	0.09	0.05	0.14
34Arts3A.txt	0.58	0.19	0.23	0.42	0.71	5.38	0.77	0.44	0.42	0.12	0.04	0.17
35Arts3A.txt	0.47	0.16	0.15	0.32	0.79	5.52	0.7	0.57	0.56	0.11	0.05	0.15
36ArtsS3A.tx	0.46	0.22	0.26	0.28	0.68	5.72	0.53	0.54	0.56	0.12	0.05	0.17
37Arts3A.txt	0.48	0.2	0.18	0.15	0.75	4.75	0.57	0.54	0.49	0.12	0.07	0.19
38Arts3A.txt	0.49	0.21	0.18	0.73	0.72	5.61	0.76	0.55	0.47	0.1	0.06	0.16
39Arts3A.txt	0.53	0.15	0.17	0.55	0.74	4.95	0.65	0.46	0.41	0.1	0.05	0.16
40Arts3A.txt	0.56	0.24	0.25	0.34	0.74	5.92	0.66	0.56	0.59	0.13	0.07	0.2
41Arts3A.txt	0.48	0.15	0.16	0.1	0.68	4.93	0.55	0.5	0.44	0.08	0.1	0.18
42Arts3A.txt	0.49	0.15	0.18	0.56	0.72	5.82	0.68	0.51	0.45	0.09	0.04	0.14
43Arts3A.txt	0.48	0.15	0.15	0.31	0.59	3.75	0.67	0.39	0.3	0.1	0.03	0.13
44Arts3A.txt	0.48	0.24	0.32	0.96	0.73	5.76	0.54	0.46	0.46	0.1	0.04	0.14
45Arts3A.txt	0.49	0.25	0.22	0.26	0.62	3.92	0.48	0.44	0.4	0.11	0.03	0.14
46Arts3A.txt	0.54	0.24	0.29	0.67	0.78	6.38	0.58	0.59	0.59	0.13	0.07	0.19
47Arts3A.txt	0.44	0.18	0.17	0.44	0.69	5.14	0.38	0.42	0.39	0.07	0.05	0.12
48Arts3A.txt	0.48	0.21	0.13	0.15	0.65	4.73	0.55	0.55	0.49	0.12	0.06	0.18
49Arts3A.txt	0.6	0.16	0.2	0.34	0.71	4.51	0.71	0.52	0.47	0.16	0.06	0.23
50Arts3A.txt	0.49	0.18	0.14	0.53	0.71	5.13	0.83	0.56	0.45	0.11	0.1	0.21
51Arts3A.txt	0.52	0.21	0.25	0.28	0.69	5.64	0.46	0.45	0.45	0.11	0.04	0.15
52Arts3A.txt	0.47	0.25	0.24	0.98	0.81	7.24	0.87	0.74	0.67	0.11	0.05	0.17
53Arts3A.txt	0.48	0.17	0.24	0.78	0.73	5.29	0.57	0.53	0.47	0.12	0.06	0.18
54Arts3A.txt	0.47	0.14	0.17	0.4	0.72	5.19	0.66	0.51	0.45	0.15	0.05	0.2
55Arts3A.txt	0.54	0.11	0.13	0.11	0.74	5.33	0.6	0.52	0.48	0.13	0.05	0.18
56Arts3A.txt	0.51	0.26	0.22	0.58	0.75	5.41	0.49	0.46	0.43	0.06	0.06	0.12
57Arts3A.txt	0.54	0.15	0.21	0.12	0.68	4.42	0.46	0.47	0.42	0.09	0.06	0.15
Average	**0.50**	**0.19**	**0.20**	**0.42**	**0.71**	**5.23**	**0.61**	**0.51**	**0.47**	**0.11**	**0.06**	**0.17**

Table 9.7 shows the calculations of 12 measures that have been applied to the sample of 25 argumentative essays written by the third year students/ school of Arts, University of Dohuk (for abbreviations of the measures, refer to the description following table 9.6).

To make a comparison between tables 9.6 and 9.7, one can compare the average scores in all the columns (see also table 9.18). In the case of lexical density, the average is 0.50 for argumentative essays but only 0.45 for narrative essays. Comparing the scores of lexical density in the two columns reveals that the students have written lexically more dense argumentative essays than narrative essays. This is because with the argumentative essays the lexical density ranges between 0.4 and 0.6 with eight scores in between 0.50–0.59 but with the narrative essays it ranges between 0.37 and 0.53 with only two scores between 0.50–0.59. In regard to the lexical sophistications 1 and 2, the average for argumentative essays is 0.19 and 0.20 respectively and for narrative essays is 0.17 and 0.19 respectively. Again with lexical sophistication, the students who wrote argumentative essays seem to have used more infrequent lexical words than those who wrote narrative essays, though not to a great extent. As for the use of infrequent verbs, despite the fact the highest score is achieved by the students who wrote argumentative essays (0.98), overall the narrative essays have more advanced verbs.

In regard to the lexical diversity, the students have scored exactly the same on the MSTTR in both the argumentative and narrative essays and less in the case of argumentative essays on the measure CTTR. However, concerning the ratio of lexical word variation (LV), the students have scored higher in the case of argumentative essays. On the different word class variation, the students have scored exactly the same on the adverb and modifier variation in both argumentative and narrative essays, higher in the case of narrative essays on the verb and noun variation, and slightly lower in the narrative essays on the adjective variation.

Table 9.8 the results of the lexical density, diversity and sophistication of the narrative essays of 4th year students, school of Arts, University of Dohuk.

FILENAME	LD	LS1	LS2	CVS1	MSTTR	CTTR	LV	CVV1	NV	ADJV	ADVV	MODV
58Arts4N.txt	0.51	0.47	0.41	1.67	0.79	6.7	0.8	0.82	0.8	0.18	0.09	0.27
59Arts4N.txt	0.45	0.3	0.25	0.39	0.75	6.28	0.79	0.67	0.56	0.09	0.08	0.17
60Arts4N.txt	0.45	0.15	0.18	0.28	0.76	5.71	0.6	0.58	0.57	0.09	0.1	0.19
61Arts4N.txt	0.46	0.12	0.14	0.3	0.71	5.04	0.62	0.55	0.47	0.12	0.09	0.21
62Arts4N.txt	0.42	0.16	0.18	0.69	0.75	6.25	0.67	0.62	0.62	0.11	0.05	0.16
63Arts4N.txt	0.44	0.12	0.15	0.18	0.74	5.81	0.58	0.61	0.59	0.15	0.07	0.22
64Arts4N.txt	0.44	0.34	0.29	0.41	0.72	5.43	0.42	0.51	0.56	0.05	0.07	0.12
65Arts4N.txt	0.44	0.15	0.19	0.71	0.73	6.03	0.6	0.54	0.56	0.08	0.05	0.13
66Arts4N.txt	0.45	0.21	0.24	0.99	0.77	6.62	0.67	0.65	0.58	0.13	0.1	0.23
67Arts4N.txt	0.45	0.24	0.27	0.91	0.74	6.63	0.56	0.6	0.62	0.1	0.06	0.16
68Arts4N.txt	0.45	0.2	0.23	0.21	0.69	5.43	0.63	0.53	0.45	0.11	0.06	0.17
69Arts4N.txt	0.44	0.13	0.16	0.21	0.79	5.71	0.7	0.6	0.58	0.1	0.05	0.15
70Arts4N.txt	0.49	0.22	0.25	0.65	0.77	5.89	0.64	0.62	0.58	0.16	0.09	0.24
71Arts4N.txt	0.52	0.12	0.16	0.35	0.76	6.37	0.67	0.53	0.48	0.08	0.06	0.14
72Arts4N.txt	0.44	0.29	0.25	1.39	0.8	6.53	0.82	0.68	0.66	0.1	0.08	0.18
73Arts4N.txt	0.43	0.26	0.28	1.28	0.74	6.21	0.72	0.62	0.57	0.11	0.06	0.18
74Arts4N.txt	0.44	0.2	0.25	0.85	0.74	5.75	0.63	0.51	0.53	0.08	0.04	0.12
75Arts4N.txt	0.42	0.28	0.23	0.68	0.72	5.43	0.6	0.63	0.56	0.11	0.08	0.19
76Arts4N.txt	0.4	0.23	0.24	0.87	0.71	5.65	0.7	0.62	0.59	0.12	0.05	0.17
77Arts4N.txt	0.45	0.19	0.17	0.36	0.77	5.78	0.76	0.62	0.58	0.14	0.05	0.19
78Arts4N.txt	0.47	0.29	0.32	0.62	0.74	6.56	0.65	0.62	0.6	0.13	0.09	0.22
79Arts4N.txt	0.54	0.08	0.09	0.33	0.76	5.54	0.73	0.61	0.58	0.14	0.06	0.2
82Arts4N.txt	0.45	0.24	0.18	0.42	0.7	4.99	0.71	0.56	0.45	0.09	0.07	0.16
91Arts4N.txt	0.4	0.19	0.21	0.54	0.72	5.73	0.54	0.52	0.49	0.09	0.06	0.15
94Arts4N□tx	0.49	0.31	0.29	0.38	0.72	5.78	0.54	0.58	0.62	0.1	0.05	0.15
Average	**0.45**	**0.22**	**0.22**	**0.63**	**0.74**	**5.91**	**0.65**	**0.6**	**0.57**	**0.11**	**0.07**	**0.18**

Table 9.8 shows the calculations of 12 measures that have been applied to the sample of 25 narrative essays written by the fourth year students/ school of Arts, University of Dohuk (for abbreviations of the measures refer to the description following table 9.6).

Table 9.9 the results of the lexical density, diversity and sophistication of the argumentative essays of 4th year students, school of Arts, University of Dohuk.

FILENAME	LD	LS1	LS2	CVS1	MSTTR	CTTR	LV	CVV1	NV	ADJV	ADVV	MODV
80Arts4A.txt	0.53	0.19	0.21	0.68	0.76	6.1	0.64	0.58	0.57	0.11	0.08	0.19
81Arts4A.txt	0.54	0.16	0.24	0.43	0.66	4.44	0.46	0.36	0.32	0.06	0.02	0.08
83Arts4A.txt	0.47	0.16	0.21	0.6	0.74	5.35	0.6	0.49	0.39	0.07	0.06	0.13
84Arts4A.txt	0.53	0.29	0.27	0.36	0.82	7.12	0.66	0.6	0.55	0.14	0.07	0.22
85Arts4A.txt	0.49	0.2	0.16	0.51	0.69	4.39	0.45	0.35	0.33	0.07	0.02	0.09
86Arts4A.txt	0.47	0.18	0.2	0.54	0.76	5.72	0.72	0.64	0.55	0.11	0.08	0.2
87Arts4A□txt	0.47	0.2	0.2	0.66	0.75	6.11	0.77	0.56	0.46	0.1	0.1	0.19
88Arts4A□txt	0.47	0.13	0.17	0.22	0.71	5.09	0.6	0.51	0.52	0.09	0.04	0.13
89Arts4A.txt	0.47	0.17	0.12	0.3	0.72	4.77	0.64	0.64	0.64	0.1	0.08	0.18
90Arts4A.txt	0.48	0.14	0.15	0.51	0.75	5.22	0.68	0.6	0.62	0.12	0.07	0.19
92Arts4A.txt	0.47	0.15	0.14	0.15	0.68	3.93	0.59	0.44	0.39	0.07	0.05	0.12
93Arts4A.txt	0.46	0.23	0.18	0.54	0.69	4.1	0.59	0.5	0.47	0.1	0.05	0.16
95Arts4A.txt	0.49	0.24	0.22	0.61	0.72	6.12	0.73	0.63	0.6	0.15	0.08	0.22
96Arts4A.txt	0.5	0.2	0.19	0.23	0.77	5.94	0.63	0.62	0.54	0.15	0.08	0.23
97Arts4A.txt	0.49	0.27	0.22	0.28	0.65	4.35	0.64	0.46	0.37	0.12	0.06	0.18
98Arts4A.txt	0.53	0.37	0.24	0.69	0.83	5.86	0.85	0.71	0.64	0.08	0.09	0.16
99Arts4A.txt	0.49	0.15	0.17	0.3	0.76	5.89	0.53	0.56	0.6	0.15	0.05	0.2
100Arts4A.tx	0.49	0.32	0.24	0.33	0.77	6.03	0.68	0.63	0.6	0.13	0.07	0.2
101Arts4A.tx	0.45	0.18	0.12	0.25	0.67	4.14	0.58	0.46	0.41	0.09	0.04	0.13
102Arts4A.tx	0.45	0.18	0.25	0.23	0.73	5.15	0.5	0.58	0.49	0.14	0.11	0.25
103Arts4A.tx	0.46	0.15	0.12	0	0.73	5.24	0.71	0.62	0.56	0.12	0.11	0.23
104Arts4A.tx	0.43	0.28	0.2	0.54	0.71	4.53	0.53	0.46	0.43	0.05	0.05	0.11
105Arts4A.tx	0.51	0.24	0.18	0.12	0.73	5.22	0.68	0.6	0.52	0.11	0.07	0.18
106Arts4A.tx	0.51	0.27	0.18	0.5	0.76	5.53	0.75	0.61	0.54	0.12	0.1	0.22
107Arts4A.tx	0.52	0.18	0.16	0.18	0.76	4.83	0.42	0.44	0.41	0.08	0.08	0.16
Average	**0.49**	**0.21**	**0.19**	**0.39**	**0.73**	**5.25**	**0.63**	**0.55**	**0.50**	**0.11**	**0.07**	**0.17**

Table 9.9 shows the calculations of 12 measures that have been applied to the sample of 25 argumentative essays written by the fourth year students/ school of Arts, University of Dohuk (for abbreviations of the measures, refer to the description following table 9.6).

As with the tables 9.6 and 9.7, tables 9.8 and 9.9 show that the argumentative essays have a higher lexical density than the narrative essays. This is not only obvious on average but on the individual level as well. Eight students have scored between 0.50–0.59 in the case of argumentative essays but only three in the case of narrative essays have scored within this range. However, with the lexical sophistication (1 and 2), the narrative essays have scored higher than the argumentative essays, meaning the students have on average used more advanced or infrequent words (words that are not within the first frequent 2000 words in the BNC). As is the case with the third year students, the ratio of the use of advanced verbs is much higher in the case of narrative essays showing that events have been described using a higher number of infrequent verbs. With regard to the lexical diversity, the students who have written narrative essays have scored higher than those who wrote argumentative essay on both the MSTTR and CTTR (0.74, 5.91 respectively compared to 0.73, 5.25 respectively). On the

level of variation in the lexical words (the LV measure), again the narrative essays have a higher average score than the argumentative essays. With word class diversity, students have used more diverse verbs, nouns and modifiers in the narrative essays than in the argumentative essays. On the level of adjectives and adverb, however, the two groups have scored the same.

Table 9.10 the results of the lexical density, diversity and sophistication of the narrative essays of 3rd year students, school of Basic Education, University of Dohuk

FILENAME	LD	LS1	LS2	CVS1	MSTTR	CTTR	LV	CVV1	NV	ADJV	ADVV	MODV
113Basic3N.t	0.45	0.1	0.1	0.39	0.74	5.03	0.6	0.56	0.56	0.06	0.08	0.14
114Basic3N.t	0.46	0.3	0.2	0.24	0.78	5.3	0.6	0.66	0.63	0.16	0.09	0.25
115Basic3N.t	0.46	0.2	0.2	0.35	0.68	5.08	0.7	0.56	0.43	0.15	0.1	0.25
116Basic3N.t	0.47	0.2	0.2	0.63	0.73	6.18	0.7	0.73	0.74	0.1	0.11	0.21
117Basic3N.t	0.47	0.2	0.2	0.53	0.7	5.49	0.7	0.59	0.5	0.15	0.08	0.23
118Basic3N.t	0.52	0.3	0.2	0.41	0.78	6.11	0.6	0.63	0.62	0.08	0.09	0.17
119Basic3N.t	0.45	0.2	0.2	0.27	0.69	4.56	0.7	0.61	0.43	0.15	0.09	0.25
120Basic3N.t	0.45	0.1	0.2	0.22	0.67	4.81	0.6	0.6	0.62	0.09	0.1	0.19
121Basic3N.t	0.48	0.3	0.3	0.4	0.74	5.31	0.4	0.46	0.52	0.09	0.06	0.15
122Basic3N.t	0.47	0.3	0.3	0.93	0.77	7.81	0.7	0.76	0.75	0.13	0.13	0.26
123Basic3N.t	0.46	0.1	0.1	0.24	0.69	5.11	0.4	0.51	0.49	0.1	0.07	0.16
124Basic3N.t	0.56	0.3	0.2	0.24	0.72	4.42	0.6	0.48	0.52	0.05	0.07	0.11
125Basic3N.t	0.44	0.2	0.2	0.25	0.73	5.94	0.6	0.61	0.59	0.1	0.07	0.17
126Basic3N.t	0.38	0.2	0.2	0.22	0.65	4.19	0.6	0.49	0.51	0.06	0.06	0.12
127Basic3N.t	0.39	0.1	0.1	0.27	0.74	5.34	0.8	0.74	0.8	0.13	0.11	0.24
128Basic3N.t	0.45	0.2	0.2	0.11	0.63	4.18	0.5	0.53	0.54	0.08	0.04	0.12
129Basic3N.t	0.49	0.3	0.2	0.15	0.73	4.72	0.7	0.65	0.62	0.11	0.09	0.2
130Basic3N.t	0.47	0.3	0.3	0.77	0.76	6.43	0.6	0.62	0.72	0.09	0.04	0.13
131Basic3N.t	0.42	0.2	0.2	0.67	0.69	5.37	0.8	0.68	0.64	0.13	0.08	0.21
132Basic3N.t	0.5	0.2	0.2	0.46	0.76	4.8	0.6	0.56	0.62	0.12	0.08	0.2
133Basic3N.t	0.45	0.3	0.2	0.62	0.73	4.86	0.7	0.67	0.65	0.09	0.09	0.19
134Basic3N.t	0.5	0.2	0.3	0.37	0.76	6.38	0.6	0.63	0.62	0.13	0.07	0.2
135Basic3N.t	0.48	0.2	0.2	0.16	0.72	4.95	0.5	0.6	0.62	0.12	0.06	0.19
136Basic3N.t	0.44	0.2	0.2	0.54	0.74	5.46	0.6	0.57	0.6	0.11	0.07	0.17
137Basic3N.t	0.43	0.2	0.2	0.71	0.7	4.62	0.7	0.65	0.63	0.18	0.05	0.23
Average	**0.46**	**0.2**	**0.2**	**0.41**	**0.72**	**5.3**	**0.6**	**0.61**	**0.6**	**0.11**	**0.08**	**0.19**

Table 9.10 shows the calculations of 12 measures that have been applied to the sample of 25 narrative essays written by the third year students/ school of Basic Education, University of Dohuk (for abbreviations of the measures, refer to the description following table 9.6).

Table 9.11 the results of the lexical density, diversity and sophistication of the argumentative essays of 3rd year students, school of Basic Education, University of Dohuk.

FILENAME	LD	LS1	LS2	CVS1	MSTTR	CTTR	LV	CVV1	NV	ADJV	ADVV	MODV
191Basic3A.t	0.52	0.23	0.23	0.39	0.74	4.88	0.38	0.43	0.42	0.09	0.07	0.16
192Basic3A.t	0.53	0.17	0.18	0.32	0.7	4.29	0.79	0.46	0.37	0.1	0.04	0.14
193Basic3A.t	0.52	0.26	0.29	0.77	0.72	5.29	0.52	0.47	0.48	0.09	0.04	0.13
195Basic3A.t	0.47	0.21	0.26	0.83	0.7	5.03	0.51	0.44	0.48	0.13	0.03	0.16
196Basic3A.t	0.51	0.23	0.2	0.12	0.75	5.03	0.51	0.5	0.42	0.12	0.06	0.18
197Basic3A.t	0.52	0.38	0.44	1.42	0.69	6.49	0.5	0.53	0.62	0.11	0.05	0.16
198Basic3A.t	0.53	0.35	0.27	0.59	0.8	6.2	0.65	0.62	0.57	0.1	0.07	0.17
199Basic3A.t	0.51	0.36	0.28	0.36	0.74	4.89	0.39	0.41	0.42	0.1	0.06	0.16
200Basic3A.t	0.47	0.11	0.17	0.11	0.7	4.7	0.51	0.45	0.39	0.12	0.06	0.17
201Basic3A.t	0.51	0.15	0.2	0.33	0.74	6.05	0.56	0.52	0.45	0.1	0.07	0.17
202Basic3A.t	0.54	0.25	0.26	0.93	0.79	5.5	0.76	0.67	0.63	0.14	0.08	0.22
203Basic3A.t	0.52	0.26	0.25	0.47	0.7	4.87	0.56	0.49	0.48	0.11	0.07	0.18
204Basic3A.t	0.57	0.2	0.21	0.14	0.74	5.33	0.64	0.6	0.55	0.13	0.09	0.22
207Basic3A.t	0.53	0.14	0.27	0.32	0.66	4.8	0.44	0.4	0.37	0.06	0.04	0.1
208Basic3A.t	0.53	0.2	0.2	0.69	0.78	4.81	0.58	0.56	0.48	0.12	0.08	0.2
209Basic3A.t	0.46	0.27	0.26	0.83	0.74	6.4	0.73	0.57	0.52	0.1	0.07	0.17
210Basic3A.t	0.5	0.23	0.27	1.15	0.73	5.51	0.61	0.48	0.45	0.13	0.05	0.17
211Basic3A.t	0.47	0.23	0.25	0.12	0.61	4.18	0.51	0.48	0.47	0.1	0.05	0.15
212Basic3A.t	0.47	0.2	0.17	0.48	0.68	4.6	0.55	0.53	0.51	0.07	0.08	0.14
213Basic3A.t	0.47	0.16	0.17	0.17	0.74	6.13	0.59	0.58	0.55	0.13	0.09	0.22
214Basic3A.t	0.5	0.19	0.19	0	0.75	5.17	0.48	0.51	0.5	0.11	0.09	0.2
215Basic3A.t	0.49	0.27	0.22	0.53	0.75	5.31	0.71	0.61	0.5	0.12	0.07	0.19
217Basic3A.t	0.46	0.21	0.21	0.37	0.59	3.47	0.67	0.34	0.25	0.09	0.05	0.15
218Basic3A.t	0.5	0.2	0.22	0.28	0.75	5.46	0.49	0.48	0.5	0.11	0.03	0.14
219Basic3A.t	0.51	0.18	0.14	0.1	0.73	4.7	0.53	0.44	0.42	0.07	0.07	0.14
Average	**0.50**	**0.23**	**0.23**	**0.47**	**0.72**	**5.16**	**0.57**	**0.50**	**0.47**	**0.11**	**0.06**	**0.17**

Table 9.11 shows the calculations of 12 measures that have been applied to the sample of 25 argumentative essays written by the third year students/ school of Basic Education, University of Dohuk (for abbreviations of the measures, refer to the description following table 9.6).

As is the case with the essays written by the third and fourth year students at the school of Arts, the written language of the argumentative essays by the third year students, school of Basic Education is also more dense than the language of the narrative essays. On lexical sophistication 1 and 2, the argumentative essays have, also, scored higher than the narrative essays. Unlike the other two cases of the third and fourth year students of the school of Arts, the argumentative essays have scored higher on verb sophistication than the narrative essays. This might be an effect of some specific individual differences as there are very low scores reaching even zero (student 214) in the argumentative essays but very high ones as well reaching above 1 (students 197, 210). The scores on verb sophistication are more homogenous in the case of narrative essays, meaning that the scores are more close to each other than the scores on argumentative essays because all students have almost produced a good number of infrequent verbs in the narrative essays.

The lexical diversity in terms of the MSTTR is exactly the same but in terms of the CTTR is different where the narrative essays have scored higher. The students who have written narrative essays have also scored higher on lexical word, verb, noun, adverb and modifier diversity but the scores on adjective diversity are exactly the same.

Table 9.12 the results of the lexical density, diversity and sophistication of the narrative essays of 4th year students, school of Basic Education, University of Dohuk

FILENAME	LD	LS1	LS2	CVS1	MSTTR	CTTR	LV	CVV1	NV	ADJV	ADVV	MODV
139Basic4N.t	0.41	0.24	0.22	0.43	0.75	5.4	0.58	0.58	0.61	0.09	0.07	0.16
140Basic4N.t	0.5	0.17	0.19	0.29	0.75	6.63	0.65	0.64	0.64	0.11	0.11	0.22
141Basic4N.t	0.44	0.17	0.19	0.58	0.69	4.81	0.53	0.49	0.43	0.1	0.07	0.16
142Basic4N.t	0.55	0.18	0.16	0.12	0.72	4.03	0.44	0.42	0.38	0.09	0.05	0.13
143Basic4N.t	0.4	0.14	0.15	0.47	0.75	5.52	0.61	0.62	0.59	0.11	0.09	0.2
144Basic4N.t	0.46	0.17	0.2	0.1	0.75	5.46	0.56	0.58	0.6	0.07	0.09	0.16
145Basic4N.t	0.47	0.17	0.13	0.14	0.77	5.63	0.84	0.78	0.73	0.08	0.05	0.14
146Basic4N.t	0.48	0.13	0.14	0.14	0.72	4.93	0.63	0.56	0.53	0.1	0.08	0.18
147Basic4N.t	0.42	0.15	0.21	0.86	0.69	4.65	0.65	0.56	0.46	0.1	0.08	0.18
148Basic4N.t	0.45	0.18	0.2	0.3	0.7	5.29	0.58	0.57	0.56	0.1	0.12	0.22
149Basic4N.t	0.46	0.21	0.2	0.24	0.75	6.09	0.53	0.57	0.55	0.13	0.09	0.21
150Basic4N.t	0.45	0.12	0.15	0.38	0.74	5.67	0.58	0.61	0.6	0.14	0.09	0.22
151Basic4N.t	0.45	0.22	0.17	0.26	0.74	5.39	0.67	0.69	0.64	0.11	0.08	0.19
152Basic4N.t	0.5	0.14	0.16	0.37	0.72	5.84	0.55	0.59	0.57	0.13	0.06	0.19
153Basic4N.t	0.49	0.09	0.12	0.11	0.72	4.26	0.39	0.44	0.46	0.11	0.04	0.15
154BASIC4N.t	0.45	0.16	0.17	0.4	0.73	5.1	0.43	0.46	0.46	0.09	0.06	0.15
155Basic4N.t	0.51	0.17	0.2	0.49	0.78	5.88	0.82	0.66	0.57	0.13	0.1	0.23
156Basic4N.t	0.48	0.19	0.24	0.58	0.74	5.28	0.53	0.47	0.47	0.08	0.06	0.14
157Basic4N.t	0.46	0.09	0.13	0.42	0.67	4.42	0.73	0.52	0.45	0.09	0.08	0.17
158Basic4N.t	0.47	0.32	0.29	0.73	0.79	7.04	0.52	0.63	0.69	0.08	0.07	0.15
159Basic4N.t	0.44	0.25	0.22	0.61	0.77	6.07	0.65	0.68	0.71	0.09	0.08	0.18
160Basic4N.t	0.48	0.29	0.26	0.43	0.76	6.3	0.63	0.6	0.5	0.1	0.09	0.19
163Basic4N.t	0.47	0.25	0.17	0.22	0.72	4.55	0.57	0.45	0.4	0.1	0.06	0.17
164Basic4N.t	0.45	0.25	0.24	0.33	0.72	5.68	0.45	0.56	0.65	0.1	0.08	0.17
165Basic4N.t	0.44	0.26	0.24	0.19	0.72	5.68	0.74	0.63	0.58	0.09	0.07	0.16
Average	**0.46**	**0.19**	**0.19**	**0.37**	**0.73**	**5.42**	**0.59**	**0.57**	**0.55**	**0.10**	**0.08**	**0.18**

Table 9.12 shows the calculations of 12 measures that have been applied to the sample of 25 narrative essays written by the fourth year students/ school of Basic Education, University of Dohuk (for abbreviations of the measures, refer to the description following table 9. 6)

Table 9.13 the results of the lexical density, diversity and sophistication of the argumentative essays of 4th year students, school of Basic Education, University of Dohuk

FILENAME	LD	LS1	LS2	CVS1	MSTTR	CTTR	LV	CVV1	NV	ADJV	ADVV	MODV
161Basic4A.tx	0.55	0.26	0.26	0.12	0.72	4.78	0.5	0.46	0.43	0.16	0.05	0.21
162Basic4A.tx	0.49	0.19	0.16	0.27	0.72	4.48	0.67	0.51	0.44	0.1	0.1	0.19
167Basic4A.tx	0.41	0.22	0.22	0.45	0.69	4.73	0.67	0.6	0.59	0.1	0.06	0.16
168Basic4A.tx	0.56	0.14	0.22	0.26	0.75	5.21	0.45	0.47	0.51	0.09	0.04	0.13
169Basic4A.tx	0.48	0.26	0.23	0.79	0.64	4.1	0.37	0.35	0.34	0.08	0.04	0.12
171Basic4A.tx	0.48	0.15	0.18	0.48	0.76	5.43	0.56	0.56	0.52	0.08	0.06	0.14
172Basic4A.tx	0.52	0.19	0.12	0	0.72	4.28	0.5	0.42	0.38	0.1	0.07	0.17
173Basic4A.tx	0.56	0.26	0.26	0.53	0.72	5.12	0.66	0.53	0.49	0.16	0.06	0.22
174Basic4A.tx	0.46	0.25	0.23	0.67	0.69	5.26	0.6	0.59	0.54	0.13	0.06	0.18
175Basic4A.tx	0.51	0.26	0.27	0.93	0.73	5.53	0.66	0.56	0.52	0.12	0.04	0.16
176Basic4A.tx	0.49	0.17	0.24	0.65	0.72	4.7	0.56	0.46	0.37	0.12	0.07	0.19
177Basic4A.tx	0.47	0.19	0.21	0.37	0.72	5.01	0.53	0.52	0.51	0.07	0.06	0.14
178Basic4A.tx	0.45	0.26	0.23	0.45	0.62	3.59	0.46	0.34	0.27	0.09	0.05	0.14
179Basic4A.tx	0.55	0.29	0.22	0.45	0.74	5.9	0.64	0.58	0.56	0.15	0.07	0.22
180Basic4A.tx	0.51	0.14	0.2	0.28	0.7	4.65	0.46	0.39	0.37	0.1	0.05	0.15
181Basic4A.tx	0.5	0.1	0.16	0.24	0.69	4.32	0.59	0.46	0.44	0.09	0.06	0.15
182Basic4A.tx	0.52	0.28	0.24	0.36	0.73	6.15	0.64	0.5	0.5	0.1	0.06	0.16
183Basic4A.tx	0.51	0.18	0.26	0.59	0.7	4.92	0.42	0.38	0.37	0.07	0.05	0.12
184Basic4A.tx	0.55	0.25	0.2	0.4	0.68	4.6	0.79	0.49	0.46	0.11	0.04	0.15
185Basic4A.tx	0.46	0.21	0.12	0.1	0.72	5.16	0.57	0.5	0.39	0.11	0.06	0.17
186Basic4A.tx	0.57	0.14	0.16	0.1	0.71	5.03	0.38	0.45	0.46	0.1	0.07	0.17
187Basic4A.tx	0.55	0.3	0.29	1.19	0.74	5.18	0.63	0.51	0.47	0.13	0.07	0.2
188Basic4A.tx	0.51	0.22	0.2	0.38	0.7	5.32	0.72	0.5	0.44	0.12	0.05	0.17
189Basic4A.tx	0.56	0.24	0.27	0.34	0.72	5.38	0.55	0.48	0.43	0.1	0.06	0.16
190Basic4A.tx	0.52	0.22	0.21	0.57	0.74	5.3	0.76	0.62	0.62	0.11	0.11	0.22
Average	**0.51**	**0.21**	**0.21**	**0.44**	**0.71**	**4.97**	**0.57**	**0.49**	**0.46**	**0.11**	**0.06**	**0.17**

Table 9.13 shows the calculations of 12 measures that have been applied to the sample of 25 argumentative essays written by the fourth year students/ school of Basic Education, University of Dohuk (for abbreviations of the measures, refer to the description following table 9.6).

The lexical density, as with all the previous cases, is higher in the case of argumentative essays, as the tables 9.12 and 9.13 show. In addition to this, the use of infrequent verbs and infrequent lexical words is also higher in the argumentative essays than in the narrative essays (notice the scores on LS1, LS2 and CVS1).

The scores on lexical diversity are higher in the case of narrative essays except for the measure of adjective variation where the argumentative essays have scored higher by 0.01.

Table 9.14 the results of the lexical density, diversity and sophistication of the narrative essays of 3rd year students, school of Languages, University of Zakho

FILENAME	LD	LS1	LS2	CVS1	MSTTR	CTTR	LV	CVV1	NV	ADJV	ADVV	MODV
241Languages3N.txt	0.49	0.17	0.17	0.5	0.74	5.12	0.46	0.46	0.43	0.09	0.05	0.14
242Languages3N.txt	0.48	0.24	0.25	0.58	0.74	6.06	0.56	0.54	0.51	0.12	0.08	0.2
243Languages3N.txt	0.48	0.25	0.26	0.53	0.75	5.38	0.64	0.64	0.61	0.13	0.12	0.25
244Languages3N.txt	0.45	0.23	0.22	0.48	0.69	5.38	0.69	0.56	0.47	0.12	0.08	0.2
245Languages3N.txt	0.44	0.29	0.23	0.29	0.75	5.86	0.63	0.63	0.6	0.07	0.08	0.15
246Languages3N.txt	0.42	0.22	0.2	1	0.65	5.35	0.76	0.6	0.56	0.08	0.05	0.13
247Languages3N.txt	0.48	0.2	0.22	0.58	0.75	5.9	0.59	0.62	0.67	0.13	0.09	0.22
248Languages3N.txt	0.45	0.23	0.24	1.1	0.72	5.66	0.65	0.59	0.57	0.1	0.1	0.2
249Languages3N.txt	0.43	0.19	0.16	0.15	0.66	3.98	0.59	0.58	0.57	0.08	0.09	0.17
250Languages3N.txt	0.41	0.24	0.21	0.75	0.71	5.04	0.72	0.65	0.62	0.11	0.04	0.15
251Languages3N.txt	0.43	0.22	0.18	0.24	0.58	3.36	0.44	0.34	0.28	0.08	0.06	0.13
252Languages3N.txt	0.44	0.14	0.16	0.43	0.65	4.24	0.53	0.41	0.34	0.09	0.07	0.17
253Languages3N.txt	0.47	0.21	0.2	0.24	0.77	5.38	0.66	0.7	0.72	0.14	0.1	0.24
254Languages3N.txt	0.46	0.26	0.28	1.28	0.75	5.87	0.67	0.6	0.58	0.09	0.07	0.15
255Languages3N.txt	0.47	0.15	0.18	0.26	0.68	4.36	0.46	0.4	0.31	0.07	0.07	0.15
257Languages3N.txt	0.44	0.16	0.23	0.28	0.67	5.32	0.51	0.56	0.62	0.08	0.07	0.15
258Languages3N.txt	0.48	0.13	0.21	0.29	0.76	5.28	0.58	0.51	0.43	0.12	0.08	0.19
259Languages3N.txt	0.45	0.25	0.17	0.23	0.71	5.08	0.71	0.68	0.67	0.09	0.05	0.15
260Languages3N.txt	0.46	0.13	0.21	0.55	0.72	4.84	0.6	0.51	0.57	0.08	0.05	0.13
261Languages3N.txt	0.51	0.16	0.23	0.29	0.72	5.06	0.71	0.55	0.62	0.12	0.07	0.18
Average	**0.46**	**0.20**	**0.21**	**0.50**	**0.71**	**5.13**	**0.61**	**0.56**	**0.54**	**0.10**	**0.07**	**0.17**

Table 9.14 shows the calculations of 12 measures that have been applied to the sample of 20 narrative essays written by the third year students/ school of Languages, University of Zakho (for abbreviations of the measures, refer to the description following table 9.6).

Table 9.15 the results of the lexical density, diversity and sophistication of the argumentative essays of 3rd year students, school of Languages, University of Zakho.

FILENAME	LD	LS1	LS2	CVS1	MSTTR	CTTR	LV	CVV1	NV	ADJV	ADVV	MODV
221Languages3A.t	0.47	0.29	0.21	0.38	0.7	5.47	0.62	0.54	0.48	0.11	0.06	0.16
222Languages3A.t	0.53	0.21	0.21	0.6	0.78	7	0.61	0.52	0.47	0.11	0.06	0.17
223Languages3A.t	0.46	0.12	0.15	0.11	0.68	4.26	0.53	0.37	0.32	0.07	0.05	0.12
224Languages3A.t	0.45	0.13	0.18	0.24	0.72	5.07	0.5	0.43	0.36	0.1	0.05	0.15
225Languages3A.t	0.54	0.14	0.15	0.18	0.74	4.75	0.43	0.38	0.38	0.07	0.07	0.14
226Languages3A.t	0.53	0.3	0.25	0.69	0.7	4.82	0.49	0.39	0.35	0.06	0.06	0.12
227Languages3A.t	0.46	0.19	0.17	0.26	0.72	4.93	0.44	0.43	0.41	0.07	0.06	0.13
228Languages3A.t	0.44	0.12	0.16	0.28	0.71	4.29	0.43	0.38	0.34	0.07	0.06	0.14
229Languages3A.t	0.53	0.23	0.2	0.15	0.65	4.12	0.61	0.4	0.32	0.11	0.04	0.15
230Languages3A.t	0.47	0.15	0.17	0.1	0.71	4.37	0.29	0.37	0.37	0.08	0.05	0.13
231Languages3A.t	0.54	0.18	0.21	0.49	0.75	5.2	0.71	0.49	0.45	0.11	0.06	0.17
232Languages3A.t	0.56	0.23	0.23	0	0.67	4.26	0.48	0.4	0.34	0.09	0.08	0.17
233Languages3A.t	0.48	0.23	0.16	0.55	0.67	3.44	0.46	0.3	0.29	0.07	0.04	0.1
234Languages3A.t	0.49	0.18	0.19	0.42	0.76	5.8	0.6	0.5	0.44	0.09	0.07	0.16
235Languages3A.t	0.49	0.3	0.26	0.62	0.73	6.32	0.66	0.76	0.77	0.12	0.06	0.19
236Languages3A.t	0.53	0.28	0.26	0.34	0.71	4.54	0.45	0.5	0.57	0.09	0.09	0.18
237Languages3A.t	0.51	0.35	0.3	0.6	0.72	5.1	0.64	0.67	0.65	0.1	0.07	0.17
238Languages3A.t	0.53	0.27	0.29	0.92	0.75	5.73	0.61	0.62	0.7	0.11	0.07	0.18
239Languages3A.t	0.5	0.32	0.24	0	0.69	5.32	0.65	0.65	0.57	0.16	0.08	0.24
240languages3A.tx	0.45	0.24	0.21	0.71	0.67	5.11	0.72	0.56	0.51	0.11	0.07	0.17
Averages	**0.50**	**0.22**	**0.21**	**0.38**	**0.71**	**5.00**	**0.55**	**0.48**	**0.45**	**0.10**	**0.06**	**0.16**

Table 9.15 shows the calculations of 12 measures that have been applied to the sample of 20 argumentative essays written by the third year students/ school of Languages, University of Zakho (for abbreviations of the measures, refer to the description following table 9.6).

Lexical density in the argumentative essays tends to be higher than in the narrative essays as the tables 9.14 and 9.15 show. The use of infrequent words is also higher in the argumentative essays whereas verb sophistication is considerably higher in the narrative essays than in the argumentative essays. As for lexical variation, the narrative essays have scored higher on all measures except for MSTTR and adjective variation, which are exactly the same in both cases.

Table 9.16 the results of the lexical density, diversity and sophistication of the narrative essays of 4th year students, school of Languages, University of Zakho

FILENAME	LD	LS1	LS2	CVS1	MSTTR	CTTR	LV	CVV1	NV	ADJV	ADVV	MODV
262Languages4N.txt	0.49	0.13	0.15	0.47	0.72	5.19	0.57	0.55	0.62	0.08	0.08	0.16
263Languages4N.txt	0.46	0.18	0.17	0.1	0.71	5.58	0.64	0.56	0.51	0.13	0.05	0.18
264Languages4N.txt	0.45	0.17	0.18	0.47	0.76	5.98	0.59	0.62	0.59	0.09	0.13	0.22
265Languages4N.txt	0.47	0.23	0.23	0.17	0.77	6.08	0.6	0.58	0.55	0.09	0.09	0.18
271Languages4N.txt	0.5	0.3	0.28	0.33	0.77	6.18	0.63	0.66	0.61	0.12	0.13	0.25
281Languages4N.txt	0.52	0.28	0.27	1.11	0.78	6.54	0.66	0.59	0.62	0.09	0.08	0.17
288Languages4N.txt	0.48	0.21	0.24	0.7	0.76	5.99	0.89	0.71	0.75	0.17	0.04	0.22
292Languages4N.txt	0.49	0.29	0.33	0.61	0.78	6.85	0.69	0.68	0.67	0.15	0.12	0.27
293Languages4N.txt	0.45	0.12	0.16	0.48	0.76	5.6	0.61	0.54	0.48	0.11	0.12	0.22
294Languages4N.txt	0.45	0.18	0.21	0.69	0.71	5.32	0.58	0.53	0.4	0.07	0.08	0.16
295Languages4N.txt	0.51	0.3	0.29	0.65	0.76	6.57	0.67	0.66	0.67	0.09	0.05	0.14
296Languages4N.txt	0.47	0.17	0.2	0.3	0.73	4.98	0.61	0.57	0.52	0.1	0.08	0.18
297Languages4N.txt	0.45	0.23	0.27	1.23	0.78	7.1	0.77	0.69	0.67	0.14	0.09	0.23
298Languages4N.txt	0.42	0.22	0.23	0	0.73	4.68	0.84	0.74	0.7	0.17	0.12	0.29
299Languages4N.txt	0.44	0.15	0.18	0.48	0.7	5.02	0.6	0.52	0.46	0.08	0.06	0.14
300Languages4N.txt	0.46	0.24	0.25	0.43	0.71	5.34	0.51	0.57	0.54	0.13	0.08	0.2
301Languages4N.txt	0.41	0.12	0.14	0.42	0.75	5.53	0.73	0.65	0.62	0.12	0.09	0.22
302Languages4N.txt	0.49	0.06	0.14	0.18	0.72	5.14	0.61	0.43	0.38	0.09	0.06	0.15
303Languages4N.txt	0.45	0.18	0.22	0.51	0.71	5.89	0.57	0.6	0.6	0.11	0.06	0.16
304Languages4N.txt	0.44	0.16	0.2	0.29	0.66	4.67	0.45	0.49	0.52	0.1	0.05	0.15
Average	**0.47**	**0.20**	**0.22**	**0.48**	**0.74**	**5.71**	**0.64**	**0.60**	**0.57**	**0.11**	**0.08**	**0.19**

Table 9.16 shows the calculations of 12 measures that have been applied to the sample of 20 narrative essays written by the fourth year students/ school of Languages, University of Zakho (for abbreviations of the measures, refer to the description following table 9.6).

Table 9.17 the results of the lexical density, diversity and sophistication of the argumentative essays of 4th year students, school of Languages, University of Zakho

FILENAME	LD	LS1	LS2	CVS1	MSTTR	CTTR	LV	CVV1	NV	ADJV	ADVV	MODV
266Languages4A.txt	0.51	0.17	0.15	0.26	0.68	4.21	0.6	0.46	0.42	0.11	0.05	0.16
267Languages4A.txt	0.49	0.23	0.24	0.34	0.75	5.37	0.58	0.58	0.55	0.15	0.06	0.21
268Languages4A.txt	0.48	0.26	0.22	0.57	0.75	5.46	0.54	0.58	0.53	0.11	0.12	0.24
269Languages4A.txt	0.43	0.23	0.25	0.92	0.72	4.91	0.5	0.46	0.49	0.09	0.03	0.13
270Languages4A.txt	0.56	0.32	0.33	0.47	0.73	4.6	0.55	0.65	0.69	0.11	0.06	0.18
272Languages4A.txt	0.54	0.35	0.34	0.45	0.76	6.07	0.9	0.76	0.7	0.16	0.12	0.27
273Languages4A.txt	0.46	0.19	0.2	0.51	0.69	4.06	0.71	0.43	0.4	0.06	0.05	0.11
274Languages4A.txt	0.49	0.26	0.36	1.06	0.76	5.87	0.72	0.58	0.52	0.12	0.05	0.17
275Languages4A.txt	0.47	0.23	0.26	0.45	0.67	4.47	0.64	0.53	0.52	0.1	0.06	0.16
276Languages4A.txt	0.48	0.26	0.23	0.79	0.67	4.6	0.7	0.56	0.51	0.12	0.06	0.18
277Languages4A.txt	0.53	0.29	0.29	0.72	0.76	5.63	0.88	0.65	0.61	0.14	0.03	0.17
278Languages4A.txt	0.52	0.21	0.27	0.5	0.75	5.52	0.56	0.5	0.48	0.09	0.04	0.14
279Languages4A.txt	0.54	0.4	0.32	0.95	0.82	6.27	0.89	0.7	0.65	0.16	0.08	0.23
280Languages4A.txt	0.44	0.22	0.23	0.66	0.58	3.94	0.59	0.45	0.35	0.12	0.05	0.17
282Languages4A.txt	0.55	0.32	0.27	0.72	0.83	7	0.8	0.74	0.72	0.15	0.07	0.22
283Languages4A.txt	0.57	0.21	0.18	0.24	0.73	3.88	0.47	0.35	0.33	0.07	0.04	0.12
284Languages4A.txt	0.5	0.33	0.29	0.47	0.75	5.56	0.65	0.66	0.6	0.23	0.04	0.28
285Languages4A.txt	0.43	0.1	0.14	0.59	0.66	4.1	1	0.57	0.47	0.11	0.07	0.18
286Languages4A.txt	0.56	0.26	0.26	0.62	0.71	4.56	0.53	0.42	0.33	0.11	0.06	0.17
287Languages4A.txt	0.51	0.23	0.17	0.55	0.78	5.02	0.69	0.53	0.47	0.08	0.07	0.15
Average	**0.50**	**0.25**	**0.25**	**0.59**	**0.73**	**5.06**	**0.68**	**0.56**	**0.52**	**0.12**	**0.06**	**0.18**

Table 9.17 shows the calculations of 12 measures that have been applied to the sample of 20 argumentative essays written by the fourth year students/ school of Languages, University of Zakho (for abbreviations of the measures, refer to the description following table 9.6).

A comparison between tables 9.16 and 9.17 reveals that on the first four measures of lexical density and sophistication, the argumentative essays have scored higher than the narrative essays whereas the scores on lexical variation vary on different measures. For example, the scores on the MSTTR, CTTR, CVV1, NV, ADVV, MODV are higher in the narrative essays but the scores on LV and ADJV are higher in the case argumentative essays.

To summarize, a number of measures could very well discriminate between the two genres. They are the measures of lexical density, the measures related to verb sophistication and variety and the measures related to lexical word sophistication and noun diversity (see table 9.18 below). The argumentative essays are obviously more dense than the narrative essays. This is more probably due to the fact that there are more complex noun phrases and a fewer number of simple clauses joined by a limited number of coordinating conjunctions as in the case of narratives. This has been shown in the chapter on grammatical complexity as phrasal complexity (especially in terms of the use of prepositional phrases in the position of

postmodification) tends to be higher in the argumentative essays, and there existed a lot of coordination in the case of narratives where many simple clauses are joined together mostly by the two coordinating conjunctions *and* and *but.* Two other reasons behind the high lexical density of the argumentative essays are (1) students have discussed more in the argumentative essays, writing a lot of words, thinking that this will make their argument stronger, and this is also reflected in the measure of lexical sophistication where it is higher in most cases in the argumentative essays, and (2) argumentation, I believe, needs more content words than grammatical words in order to enrich the argumentation with more meaning. In this case when the pedagogical aim of teaching is to help students learn more lexical words, argumentative essays are necessary.

In addition to this, in more cases, the use of infrequent verbs is higher in the narrative essays and the use of varied verbs (the scores on the measure CVV1) is higher in almost all cases of the narrative essays. This might be because narration is a description of events and thus the students have used more varied verbs than in the case of argumentative essays. However, if a measure of verb density was included (number of verbs to the number of all other words), it would be a better indicator of the high number of verbs in narrative essays. Moreover, the use of variety of nouns is also high in the narrative essays. The measure of NV is higher in all cases in the narrative essays. This also has a useful implication for language teaching. Teachers can effectively teach students lexical verbs and nouns through narrative essays.

For showing whether the points discussed above can be proved right, I take two examples, one from a narrative essay and one from an argumentative essay. The following are the examples:

e.g. 9.1 S(2)

In the first moment while I *came* to college I *started feeling* rather shy and nervous. My mind *became* full of thoughts until I *was* afraid of *losing* my control all these because of some direct reasons which I *faced* at the beginning of my movement *coming* to college

e.g. 9.2 S(290)

Many developing countries or regions *use* the language of a developed country. In Iraqi Kurdistan, recently the English language *become* necessary for any one who *wants* to communicate with the outer world.

Through a close inspection of these two examples, one can see that example 9.1 (from a narrative essay) contains more verbs than example 9.2 (from an argumentative essay) and the

noun phrases in this example are not very long and complicated like the noun phrases in example 9.2.

Table 9.18 the averages of the scores in tables 9.6 to 9.17 (i.e. a summary of these tables)

Student groups	LD	LS1	LS2	CVS1	MSTTR	CTTR	LV	CVV1	NV	ADJV	ADVV	MODV
3rd year Arts (Narrative)	0.45	0.17	0.19	0.44	0.71	5.30	0.56	0.53	0.51	0.10	0.06	0.17
3rd year Arts (argumentative)	0.50	0.19	0.20	0.42	0.71	5.23	0.61	0.51	0.47	0.11	0.06	0.17
4th year Arts (Narrative)	0.45	0.22	0.22	0.63	0.74	5.91	0.65	0.60	0.57	0.11	0.07	0.18
4th year Arts (argumentative)	0.49	0.21	0.19	0.39	0.73	5.25	0.63	0.55	0.50	0.11	0.07	0.17
3rd year Basic Education (Narrative)	0.46	0.20	0.20	0.41	0.72	5.30	0.62	0.61	0.60	0.11	0.08	0.19
3rd year Basic Education (argumentative)	0.50	0.23	0.23	0.47	0.72	5.16	0.57	0.50	0.47	0.11	0.06	0.17
4th year Basic Education (Narrative)	0.46	0.19	0.19	0.37	0.73	5.42	0.59	0.57	0.55	0.10	0.08	0.18
4th year Basic Education (argumentative)	0.51	0.21	0.21	0.44	0.71	4.97	0.57	0.49	0.46	0.11	0.06	0.17
3rd year Languages (Narrative)	0.46	0.20	0.21	0.50	0.71	5.13	0.61	0.56	0.54	0.10	0.07	0.17
3rd year Languages (argumentative)	0.50	0.22	0.21	0.38	0.71	5.00	0.55	0.48	0.45	0.10	0.06	0.16
4th year Languages (Narrative)	0.47	0.20	0.22	0.48	0.74	5.71	0.64	0.60	0.57	0.11	0.08	0.19
4th year Languages (argumentative)	0.50	0.25	0.25	0.59	0.73	5.06	0.68	0.56	0.52	0.12	0.06	0.18

Table 9.18 includes all the averages from tables 9.6 to 9.17 for all the three schools for both third year and fourth year students

Before concluding this section, it is important to note that it is not easy to tell whether the results would be the same if the tool was applied to a larger number of data. The number of essays in the sub corpora used in this chapter depended on the point that all the sub corpora must have an equal number of essays.

9.4.3 A comparison between the lexical complexity of the third year students' essays and the lexical complexity of the fourth year students' essays

The figures in table 9.18 can be compared in terms of years of study to show how lexical complexity differs from one year to the other.

To compare the narrative essays written by the third and fourth year students of the school of Arts, we have to compare the averages in the table 9.18. As this table shows, the third and

fourth year students' essays (at the school of Arts) have scored exactly the same on lexical density (0.45) but the fourth year students have scored higher on all the measures of lexical sophistication including the CVS1 and lexical variation as well. This demonstrates that there is a development in lexical variation and sophistication from the third year to the fourth year.

Now comparing the performance on the argumentative essays (school of Arts), it is clear that the third year students have scored higher by only 0.01 on lexical density and have scored high on lexical sophistication-2 and CVS1 but a little lower on lexical sophistication-1. In regard to lexical variation, the fourth year students have scored higher on all measures except for adjective and modifier variation on which they scored exactly the same as third year students. This also shows that there is a little development from year 3 to year 4 in terms of lexical variation but not lexical density and sophistication.

Considering the performance of the third and fourth year students at the school of Basic Education on narrative essays, they have scored the same on lexical density (0.46). On lexical sophistication, the third year students have scored the same on both LS1 and LS2 (0.2) but higher than the fourth year students and they have also scored higher on the use of advanced or infrequent verbs (the measure CVS1). However, the measures of lexical variation have yielded different results. While the TTR ratios are higher, though not to a great extent, in the case of fourth year students, on all the remaining lexical variation measures, the third year students have scored higher (except on the adverb variation, which is 0.08 in both cases). This shows that there is not a very noticeable vocabulary development from year 3 to year 4.

The third and fourth year students at the Basic Education school have both scored high on lexical density in the argumentative essays. The fourth year students have, however, scored higher only by 0.01. As for performance in terms of lexical sophistication and verb sophistication, the third year students have scored higher. The results on the lexical diversity measures vary. No development can be seen in the results of the following measures: LV, ADJV, ADVV, and MODV. Nevertheless, the third year students have scored higher on the MSTTR, CTTR, CVV1 and NV measures.

There is a small difference between the lexical density of the narrative essays written by the third and fourth year students at the school of Languages, University of Zakho with the fourth year students scoring higher only by 0.01. On the lexical sophistication measures, the third year students and fourth year students have scored the same on the LS1. However, the LS2 is higher in the case of fourth year students and the CVS1 is higher in the case of third year students. On the lexical variation measures, the fourth year students have scored higher than

the third year students on all the measures, meaning they have written essays with more varied words and there is a clear development from year 3 to year 4 in the use of different words.

The third year and fourth year students' (school of Languages/ University of Zakho) performance on argumentative essays shows that the fourth year students have scored higher on all the lexical complexity measures except for lexical density and one of the lexical variation measures, namely the ADVV, where they scored the same as the third year students (0.50, 0.06 respectively).

In all these comparisons, one can say that there is development in the use of varied vocabulary. More fourth year students have tried to use different words in their essays.

9.5 Conclusion

This chapter has dealt with lexical complexity. Reviewing the literature on lexical complexity, it became evident that it has three dimensions namely, lexical density, lexical diversity and lexical sophistication. Read (2000) has added another dimension to the list, which is the proportion of errors. Researchers have operationalized these dimensions through the use of various measures. A number of these measures have not taken the sample length into account, like the type/token ratio (TTR) as a measure of lexical diversity and the calculation of the number of different words (NDW). That is why researchers have found other alternatives for these measures like the Corrected TTR (CTTR), the LogTTR, the Uber index etc.

Despite the fact that researchers have used various types of measures, they have not been able to use many measures together or use a number of them on large sets of data. The development of an automated application of these measures has, thus, been both important and necessary. For this purpose, the Lexical Complexity Analyzer (LCA) has been developed by Xiaofei Lu, in which the calculation of 25 measures of lexical complexity has been automatized. Thus, analysis became a matter of minutes rather than days of manual labor. As this is a newly invented tool, to test whether the tool is reliable or not, I have compared some of the calculations, such as the calculation of types and tokens to the calculations made by Antconc version 3.4.3m as this is a tool that has been used a lot by different researchers. It emerged that there was a correlation of 0.996 for types and 0.998 for tokens for a number of 25 essays chosen for the purpose.

The LCA tool is used to calculate 25 measures but only 12 of them have been selected for comparison in this study. The choice was based on a number of points like the coverage of all the three dimensions and the measures' degree of sensitivity to sample size.

The application of these automated measures to 280 essays of the data yielded different results. A comparison between the argumentative and narrative essays has shown that the argumentative essays are more dense than the narrative essays which might be due to the use of a high number of simple clauses coordinated by a number of coordinating conjunctions like *and* and *but* in the narrative essays and the use of long and complex noun phrases in the argumentative essays. It is also concluded that the verb variation is higher in the narrative essays and this can be expected because narration is a description of events.

The comparison between the results of the third year students and fourth year students has revealed mixed results. In addition to the differences in the results of individual measures, in some cases the third year students have scored higher on almost all measures like the case with the performance of the 3rd and 4th year students at the school of Basic Education on narrative essays, but in more cases, the fourth year students have scored higher on the majority of measures e.g. the fourth year students' (of the school of Arts) performance on narrative essays, the fourth year students' (of the school of Languages) performance on argumentative essays. This tells us that one can detect a development in vocabulary from year three to year four unlike the case with grammatical measures. This might probably be due to the fact that students pay more attention to learning vocabulary without learning the syntax of placing these vocabularies together in well-formed structures.

Before bringing this chapter to an end, I think it is necessary to consider the strong points and limitations of the LCA and what can be done to decrease these limitations and make the analysis better. The main advantage of the use of the LCA is that it saves a lot of time and efforts and it gives the researcher the benefit of measuring the three dimensions of lexical complexity in a matter of minutes. The other good feature of LCA is that it includes measures that are sensitive to the length of writing e.g. the corrected TTR, the corrected VV1, which makes it possible for the researcher to use essays of any length. In addition, the LCA offers the advantage of not only measuring diversity and sophistication of all lexemes in general but goes further than this and measures the diversity and sophistication of all parts of speech e.g. the measures CVS1, NV, ADVV, MODV, ADJV.

However, it is also important to state the limitations of the analysis done in this chapter. The first and major limitation is that because of the big number of the spelling errors and mistakes the LCA might not have worked well because a computer tool cannot recognize that a word like *wer* provided in the place of *were* is just a wrong variant of *were* and should not be considered as a type. In this way, the LCA program might have calculated more types than a given essay has and a student with more spelling mistakes would have been rewarded rather than penalized. This would only affect the measures involving types and tokens like the TTR. To remedy this case, it is better for the researcher to correct all the spelling errors and mistakes before using the program or do the work manually for calculating the measures that involve the types and tokens. Both of these are not very good solutions because they are very time-consuming.

Notes

1. Lexical complexity is also called ‘lexical richness’ in some publications but I will use ‘lexical complexity’ in this chapter.
2. The files have to first be zipped as a batch and then uploaded.
3. I have to thank Dr. Paul Thompson and Dr. Haiyang Ai for their kind support in teaching me how the tool works.

Chapter Ten

Conclusions, Pedagogical Implications and Future Research

This chapter presents three main points that are based on the research for this thesis: conclusions, pedagogical implications and suggestions for further research. It first discusses the main conclusions of the present study and then explains the implications for writing pedagogy and writing assessment. In section 10.4, it provides a number of topics for future research. The chapter ends with my conclusions.

10.1 Conclusions

This thesis treats the problem of how to measure the accuracy and complexity of English as a foreign language (EFL) writing especially in low level students. It is clearly easier to code native speakers' writing and advanced non-native speakers' writing than the low level writing exemplified in this thesis. Analysis might even be done automatically using computer software. In this thesis, different methods have been used for the purpose of measuring these two constructs i.e. accuracy and complexity. The major aim was to look for a way to measure accuracy and complexity positively rather than negatively. To put it more simply, I was seeking a possible method of measurement that does not simply result in taking a decision such as 'the whole essay is an error'. In other words, the aim was to find objective measures for accuracy and complexity in language. These measures should give more credit than is usual to the writers by assessing writing positively rather than negatively, but at the same time should not allow too much of the language to be correct.

The first method reported is error analysis. Errors were identified, described and categorized. The total number of errors was calculated for 48 essays and three groups of errors were identified. Although a very detailed error taxonomy was used (the Université Catholique de Louvain's taxonomy), a great deal of the language could not be analyzed. I had to add different error tags among which were 'unclear stretch', and 'ill-formed stretch'. These two labels were given to pieces of language that were so problematic that they could not be analyzed with error analysis. It was apparent that, if only error analysis was applied for measuring accuracy, all the language in such stretches would remain unanalyzed. Also, it became apparent that a number of essays included more errors than correct forms. These two cases inspired me to think of another method that could measure language more positively rather than negatively or could measure correctness rather than erroneousness. I named the

method ‘correctness analysis’. First, an existing method was tried using T-units (defined as “the minimal terminable units” or the “shortest allowable sentences”, originally identified by Hunt, 1966:737) and clauses (with their three types: finite independent, finite dependent and non-finite) as units of analysis. Three different measures were used, namely the ratio of error-free T-units, (EFT/T)[1], the ratio of error-free clauses (EFC/C)[2] and the ratio of error-free clauses to the total number of T-units (EFC/T).

Despite the fact that this method showed that the fourth year students in both schools (school of Arts and school of Basic Education) were slightly (though statistically not significant) better than the third year students, it still did not give enough credit to the students i.e. their language was not evaluated fairly. This is because the method (1) still depended on errors, (2) took as units of analysis T-units and clauses which are more prone to the occurrence of errors than smaller units like phrases or two-word units, (3) was not flexible i.e. the units to be identified had to be either T-units or clauses, and (4) did not take error severity into consideration.

As a result of these points, it was necessary to think of a new way of measuring correctness, a method that would be more flexible in regard to the choice of the basic units of analysis. I have devised another method called ‘various-units-based correctness analysis’. With this method, all correct units are considered for analysis, no matter whether they are sentences, clauses, phrases, or even two- or three-word units, and the following new measure has been developed for calculating the results of this method.

$$accuracy = \frac{[no.\,of\ correct\ sequences\ (CS) \times (mean\ length\ of\ correct\ sequences\ (\ No.\,of\ words\ in\ CS/\ NCS\ )\ ])}{total\ number\ of\ words\ in\ essay}$$

The measure is an attempt to identify what proportion of the total words in the essay are correctly used. It measures accuracy by multiplying the number of correct sequences by their average length and then dividing the result by the total number of words in the essay. The average length of correct sequences can be calculated by dividing the total number of words in the correct sequences by the total number of these correct sequences.

Although this method worked well with measuring accuracy, a number of methodological problems occurred when it was applied. For example, identifying different units was not easy and still no error severity was taken into account. This was the reason for devising another new method that could integrate all the previous methods (i.e. error analysis, T-unit- and clause-based correctness analysis and various-units-based correctness analysis) in one

method. I called the method 'An Integrated Approach to Achievement' or (IAA). With this method, most of the language is assessed whether correct or incorrect, and students are rewarded as well as penalized fairly. In other words, the method could draw a fine line between levels of achievement.

Hence, one of the main achievements of this study is the development of this method of assessing accuracy (i.e. IAA) because it seems to have a number of advantages: (1) it takes error severity into consideration, (2) it is based on various units, (3) it takes both correctness analysis and error analysis into consideration, (4) it does not leave the unanalyzable units as they are but further analyzes them into small correct and almost correct sequences, (5) it differentiates well between essays that are similar but not identical in proficiency, (6) it can be used both in language writing assessment and writing pedagogy. Although I have measured accuracy with different measures, the results of all the measures are in most cases consistent with one another i.e. the students who have performed well on one measure have performed well on the other measures as well. However, the IAA can be considered a relatively objective measure that can integrate all the methods and that can reward as well as penalize students.

As a complement to the accuracy measures, I have also applied measures of complexity to the data collected for this thesis. Two types of complexity measures have been applied: syntactic and lexical. For measuring syntactic complexity, I have used different measures to measure sentential, clausal and phrasal complexity. Phrasal complexity and complexity with non-finite clauses and phrases are regarded as higher levels of language complexity i.e. higher than coordination and subordination with devices like *because*. Lexical complexity is assessed automatically with a software package called the Lexical Complexity Analyzer. This software measures 25 measures of lexical complexity and gives a full picture of lexical density, diversity and sophistication. The fourth year students have shown a development in lexical complexity or the use of vocabulary. This case was unlike the case with the syntactic complexity where it hardly showed any development from year 3 to year 4.

Going through this conclusion and the conclusions at the end of the practical chapters, I can provide answers to the three research questions in chapter one. uestion one: Is there evidence of improvement in performance between students in year three and four (in the three schools) in terms of accuracy and complexity?, o significant difference could be between third and fourth year students in the three schools on the levels of accuracy and syntactic complexity. However, on the level of lexical complexity, there was more noticeable improvement from year three to year four in most cases. This may that students pay more attention to vocabulary

learning than syntax of how to vocabular in well-formed bigger units like clauses and sentences. uestion two: How do different measures compare? What does each measure show?, ecause of the quality of the data I analyzed in this thesis (high in ideas and low in syntax), error analysis did not prove to be a good measure. I tried adding more codes but still it did not work well. The T unit and clause-based correctness analysis was not proved to be a good measure either T units and clauses are long units and they were more prone to errors. Also, error severity was not taken into consideration. I made an attempt to take smaller units in the various-units-based correctness analysis. Although this was a better measure, it still did not fulfill the need of having a measure that is objective enough does not ence I have devised the IAA combining all these measures into one measure. As stated earlier, this method is considered to be the most objective measure among all the other measures introduced in this thesis. It measures students' performance precisely taking consideration of what is right but at the same time does not show bias towards students.

- Concerning the measures of syntactic complexity, I would support the view that measures considered separately because subsuming the whole area under one joint measure may not be a good idea. For example, measuring all the subordination and all the coordination together might not be a good idea because this obscures information about the item-based complexity. For instance, when the subordination was high, it was, in most cases, due the high use of *because* and the use of other subordinating conjunctions was not so common. High degrees of coordination do not seem to indicate, in all cases, 'better' use of language or higher levels of complexity than less use of coordination. On the contrary, the use of coordination is sometimes might only indicate that the learner has not mastered other ways of elaborating the language for communication and resorts to the use of *and*, and *but* to make his language more complex. The phrasal complexity measures seem to be able to draw a finer line between the levels of students because they were considered separately from each other. All post nominal modifiers have been used on their own and thus gave a clearer picture of the phrasal complexity. uestion number three, it became obvious that the learners' performance could be assessed positively rather than negatively but at the same time objectively without a bias in favor of the learner.

10.2 Implications for language testing

There are two drawbacks to the use of the method of IAA. Firstly, it is based on extensive manual coding of the data and is thus very time-consuming; secondly, and as a consequence,

it is not clear how it could be automated. However, if turning IAA into computer software were possible, it would be a significant contribution to the field of language testing, since automation is the greatest current growth area in language testing (Kate Knill, personal communication). My future ambition is to identify computer programmers who would be capable of developing this method into computer software for the automatic grading of written tests. If such an attempt succeeds, it will be beneficial for teachers of writing, freeing their time for other aspects of teaching. As a teacher of writing, I have experience of the difficulty of grading essays and exams, and most teachers of writing complain about this issue. This is in addition to the fact that IAA is a very objective measure and hence will assist teachers to depend less on subjectivity in holistic rating.

'Correctness analysis' in general and IAA in particular can also be important in assessing writing because instead of looking at the negative aspects of the writing (errors), the teacher looks at the positive side of the writing (correct forms). Choosing to assess from the perspective of errors may make the teacher intolerant. When I began my research, I expected that most units in the essays would be unanalyzable (that is, they would be unanalyzable T-units), but the results of IAA were not congruent with my expectations; I discovered that only a small number of the units remained unanalyzable.

I have had the opportunity to present my ideas to teachers at the department of English at Dohuk University in Iraqi Kurdistan, who helped me by recoding the essays. They responded to the new method with enthusiasm and expressed their desire to apply it in teaching writing. Another important point is that it is more encouraging for students to get feedback from the teacher telling them where they have produced correct pieces of language rather than only obtaining feedback on errors.

In assessing writing, teachers need to take complexity as well as accuracy into account, as it is known that complexity and accuracy are sometimes in a 'trade-off' relationship i.e. one of them is sacrificed at the expense of another (see chapter two, section 2.5, point number 4). However, the various components of complexity have to be considered because a student who has used a lot of subordination with, say, *because* is not at a higher level of proficiency than a student who has used fewer cases of subordination but with various types of subordinating conjunctions. In this case, complexity should be assessed in terms of quality as well as quantity.

While engaging with my data during the analysis, it became clear to me that many of its features, such as the use of fillers and sub-clausal units, were in many respects similar to those used in spoken English. Thus, the methods proposed in this research might have applications to the assessment of spoken language as well as of written language.

10.3 Implications for the teaching of writing

Polio (1997:102) highlights that although other areas than linguistic accuracy (like linguistic complexity, coherence, and content) might be more important in writing, linguistic accuracy remains an important topic in "SLA, L2 writing assessment and L2 writing pedagogy". Writing is one of the important components of language learning, and it is a skill that has to be taught well and assessed as objectively as possible by raters and assessors. As a result of this and also since the teacher acts as an assessor or rater in most cases, it is very important that teachers learn from the way students write in order to devise new techniques of teaching and assessment accordingly.

Based on the fact that "instructed SLA researchers seek to understand phenomena that make a difference in teaching and learning, first and foremost" (Norris and Ortega, 2009:557), I have found it very important to list and explain a number of the implications that this piece of research carries for language teaching and learning in general and for writing in particular.

Another argument for the importance of focusing on writing pedagogy and assessment is made by Byrnes and Manchón (2014 cited in Ruiz-Funes, 2015:3) who point out that the practice of writing assists in the process of learning a language itself. Ruiz-Funes (ibid) summarizes their argument as follows:

> writing promotes language learning and development because of the problem-solving nature of writing itself, the availability of time writing acts provide, and the attention paid by learners to language use and processing primarily in complex meaning-making tasks.

The following sub-sections are descriptions of a number of writing pedagogical implications that I hope will be of use to the teachers.

10.3.1 Pedagogical implications of the chapter on error analysis

Having the students commit fewer or no errors at all in language is the ultimate goal of all FL/SL teachers. Therefore, identifying the number and types of errors the students make plays

an important role in helping teachers plan differently for their lessons and advise curriculum designers to focus on some areas more than others. Errors, in other words, help the teacher to know the level of their students. As highlighted in chapter four, two kinds of error are committed by the students, local and global errors. The teacher can make use of both types in teaching. The local errors can help the teacher specify the problems their students are facing, and they might be treatable with teaching. For example, with errors in articles, the teacher may try to identify the cause of the errors, (which might sometimes be the system of the students' L1 or it might only be the students' paying less attention to this particular linguistic item) and thus look for ways of treating the problem. Examples such as the following are treatable in teaching, where the basic sentence structure is not so violated as to make the sentence incomprehensible. Consider the following examples; in each case the original sentence (a) is emended in (b), with the emendation shown in square brackets:

e.g. 10.1a S(14)

but I have to think of coming days and how to make My las week in college great moments for my life and future.

e.g. 10.1b

but I have to think of [the] coming days and how to make My las week in college great moments for my life and future.

e.g. 10.2a S(4)

After we got money we take him to a very good doctor in sham and after five months my father recovered and was a glad.

e.g. 10.2b

After we got money we take him to a very good doctor in sham and after five months my father recovered and was a [0] glad.

The treatment can be administered through written or verbal corrective feedback. It can also be carried out through collecting a batch of these errors and then presenting them in class anonymously along with their correct versions and explaining why such structures are erroneous. If many students are making such errors with articles, the teacher can seek a different way of teaching that may help students learn the item more easily.

However, in my analysis of errors, I found that students do not only make these simple local errors which do not cause significant problems in sentence comprehension but also make very global errors that make the language understandable only if taken in small pieces. Alternatively, the errors make the language incomprehensible even in the wider context, the

whole essay or the non-linguistic context. I have labeled these errors ill-formed stretches and unclear stretches. In the essays that contain such an erroneous type of language it is not easy for the teacher to pinpoint specific types of errors and hence bring them to the attention of the student or teach the right version of the specific erroneous linguistic item. These ill-formed stretches do not only contain one or two types of concrete errors that can be identified easily to the students but they are sentences or stretches that are composed of pieces of language (or only words) juxtaposed without reasonable grammatical relations between them. The stretch is sometimes understandable but still so ill-formed that the errors are difficult to identify and point to. Consider the following examples:

e.g. 10.3 S(270)

It be dangrous disease for health is like bad thing to life smoking make many affected up to the person who smoke cigarette, it be banned.

e.g. 10.4 S(24)

I say something about my day when I saw or met my friend and also I will descripe inner beauty and outer beauty about my friend but firstfull:- I will say about the day was a very nice day was a sunshine and beautifully day and I was a relax in day and happy so any how she came in the beautiful day.

Now looking at the first stretch what the writer might have meant is:

(1) It is a dangerous disease for health. (2) It is like a bad thing for life. (3) Smoking has many effects on the people who smoke cigarettes; (4) it must be banned.

This rewrite consists of four clauses, separated here by punctuation. However, in the original version the four clauses all occur together in one stretch. Therefore, the errors are so interconnected that they cannot be easily shown to the writer. In the second example (10.4), it is even more difficult to separate the components because there is even more violation of grammatical relations between the words. The question here is, how can we make use of such errors in teaching? Or how can we remedy such cases? The answer to this question can be made more clear in the coming sections. However, I can present a part of the answer in the following paragraph.

The students can be given back their work without any feedback and asked to explain what they actually meant by the stretch. Then, they may be asked to rewrite exactly the same ideas with simpler sentences with help from the teacher. The teacher may then compare what the students have written with their original attempts and try to point out the areas where the ideas needed to be expressed in simpler sentences connected with different connecting devices like coordination and subordination conjunctions, anaphoric and cataphoric reference etc.

Another way of benefiting from these ill-formed stretches in writing is that the teacher can present some of these stretches to the students anonymously, try to isolate the number of ideas expressed in one stretch, and ask the students to express these ideas again in smaller stretches that might be easier for the students to grasp. The teacher then can show the students how to combine the small stretches into longer ones or ask the students to do so themselves.

10.3.2 Pedagogical implications of the chapters on correctness analysis

In cases like examples 10.3 and 10.4, it is easier for the teacher to show the correct pieces of language than the erroneousness sections. It is not only easier but also more encouraging to the students. It is worth noting that the smaller the pieces of language, the easier it will be for the teacher to find more of them. That is why units like T-units, C-units, and sentences are always more prone to error than phrases and two-word units or three-word units. The teacher can, therefore, divide the text into phrases or any unit that they find correct or with only one simple error (from the various-units-based correctness analysis). The teacher can divide all the problematic stretches of language or the so-called unanalyzable T-units (see chapter 7 for an explanation of these units) into smaller units that are correct and 'almost correct' (an application of the IAA method) and give them back to the students. The students may be asked to identify the correct and incorrect sections and to correct the incorrect ones. Here the teacher would have narrowed down the domain and extent of errors from global errors to local errors where the students can find them more easily.

As an example we can identify the following units in examples 10.3 and 10.4

Example (10.3)

1. *It be dangrous disease* (incorrect) = *It* <u>is</u> <u>*a*</u> *dangerous disease* (correct) (2 errors)
2. *is like bad thing* (incorrect) = <u>It</u> *is like* <u>a</u> *bad thing* (correct) (2 errors)
3. *the person who smoke cigarette* (incorrect) = *the person who smoke<u>s</u> cigarette<u>s</u>* (correct) (2 errors)
4. *it be banned* (incorrect) = *it <u>must</u> be banned* (correct) (1 error)

Example (10.4)

5. *I say something about my day when I saw or met my friend* (incorrect)= *I* <u>will</u> *say something about* <u>the</u> *day when I saw or met my friend* (correct) (2 errors)

6. *and also I will descripe inner beauty and outer beauty* (incorrect)= *and also I will describe* the *inner and outer beauty* (correct) (1 error)
7. *my friend (correct)*
8. *I will say about the day* (incorrect) = *I will* describe *the day* (correct) (1 error)
9. *was a very nice day* (incorrect)= It *was a very nice day* (correct) (1 error)
10. *was a sunshine and beautifully day* (incorrect)= It *was* sunny *and* a beautiful *day (correct)* (4 errors)
11. *I was a relax* (incorrect)= *I was* relaxed (1 error)
12. *the beautiful day* (correct)

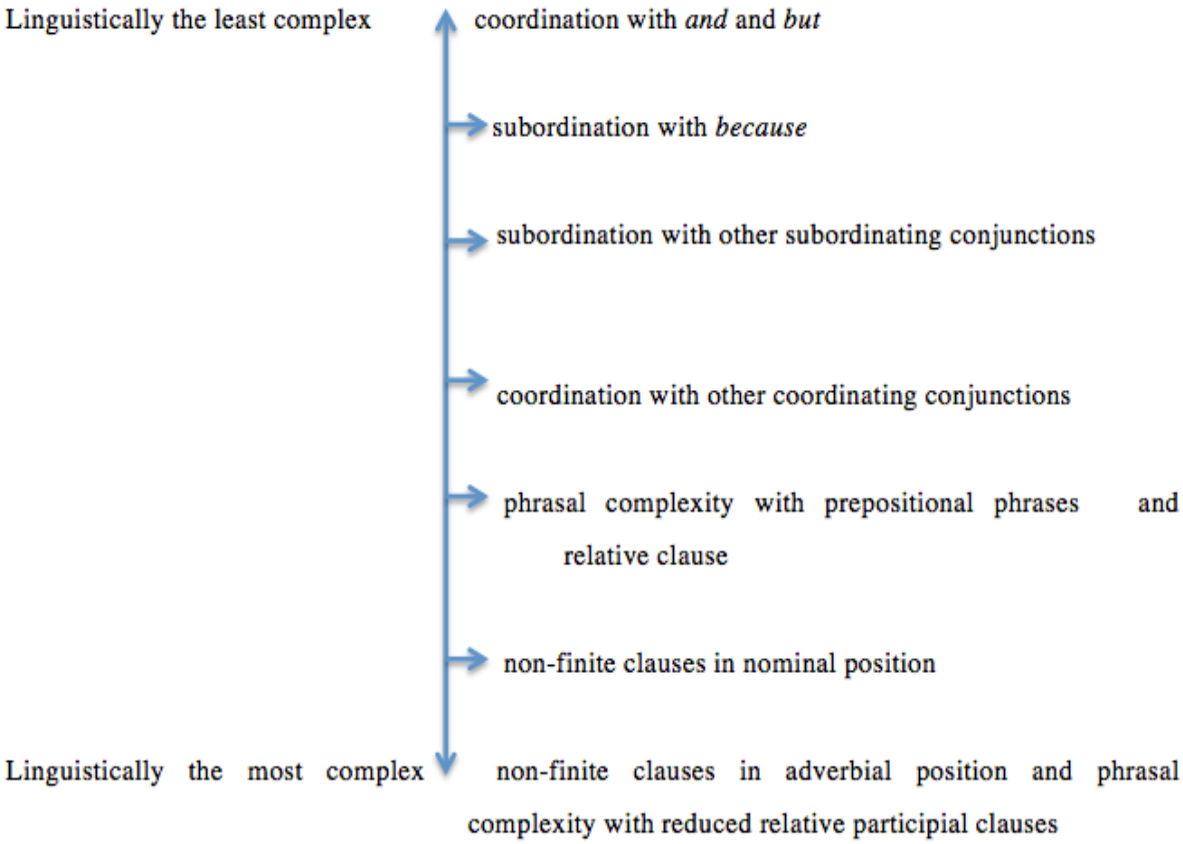
These pieces of language can be given back to the students for correction and recombining into bigger units.

10.3.3 Pedagogical implications of the chapter on syntactic complexity

The chapter on syntactic complexity (Chapter 8) has demonstrated mixed results. The most outstanding ones can be summarized in a number of points: (1) phrasal complexity is very rare in the data, especially those structures involving both relativization and reduction such as participial clauses (e.g. 10.5 the girl standing the in the corridor is my friend), (2) non-finite clauses functioning as nominals and adverbials (e.g. 10.6 S(139) *Thinking of that the next hour I have a rest, I spent the whole hour in her office* are also very rare in the data, (3) there is excessive use of coordination using coordinating conjunctions like *and* and *but* especially in the narrative data, (4) students write very long T-units especially in the argumentative essays, (5) argumentative essays have been found to be more complex even in terms of phrasal complexity, (6) a number of students have achieved a balance between complexity and accuracy and some others are good at one at the expense of the other.

These points tell us that complexity in learner language is a continuum with simple processes like coordination at one end, the use of non-finite phrases in adverbial position and phrasal complexity with reduced relative participial clauses at the other, with the other structures lying between them. An attempt to demonstrate this continuum is shown in figure 10.1:

Figure 10.1 The continuum of complexity in language pedagogy



This continuum, I believe, is very important in teaching complexity devices. It is important that curriculum designers and also teachers pay attention to such an issue. They may start teaching coordination with *and* and *but* followed by subordination with *because* and then subordination with other devices like *unless*, *since*, *as long as*. The teachers could continue with the continuum until reaching the most difficult cases (the non-finite clauses and reduced relative participial clauses). It is not necessary that all items be taught in only one year of study but they can be taught sequentially in successive years with revisions of what has been taught each year in the year that follows it.

Another significant implication of chapter eight is that argumentative essays are important for the teaching of complexity devices. This is because argumentative essays have been shown to be more complex in regard to phrasal complexity. However, narration has been shown to involve the use of a lot of coordination and hence coordination conjunctions can be taught through narrative writing. What is more important is that when the students learn how to use subordination and coordination with different conjunctions, it is better that the teachers plan lessons on argumentative essays asking the students to use coordination devices and narrative

essays asking students to use subordination devices. This will help the students learn the use of both types of conjunctions in both contexts.

10.3.4 Pedagogical implications of the chapter on lexical complexity

Three important points can be concluded as pedagogical implications for the teaching of writing from chapter nine. The first is that the argumentative essays appeared to be more dense than the narrative essays. One of the reasons could be, as stated earlier, that this type of essay has more complex noun phrases where many lexical items as opposed to grammatical items exist. There is a vital point here that teachers can make use of, which is that students can be directed to learn lexical words through asking them to write argumentative essays. The second point is that it appeared that there was development (in most cases) from year three to year four in terms of lexical density, diversity, and sophistication. This could possibly encourage teachers to teach more vocabulary and students to learn more lexical items as they proceed in their study. The third point is that students seem to have used more various infrequent and sophisticated verbs in narrative essays. In this case, narrative essays can be a good tool for teaching verbs.

Although the following views might not be direct conclusions from chapter nine, they seem to be beneficial for teachers of English. It is important for the teacher of English to take lexical density, sophistication and variation into consideration while teaching vocabulary and designing materials on vocabulary. Students cannot be taught only a large number of words, or only various words or only infrequent words but they have to be taught a large number of various frequent and infrequent lexical items. The frequency and infrequency can be taken to be based on specific teaching contexts. For example, some infrequent or sophisticated words might be complex for a number of students, yet they might not be infrequent for others in a different context. Therefore, context is very important while teaching vocabulary. Also, vocabulary may not be taught in isolation but in text in the case of writing and communicative activities like role plays, dialogues etc. in the case of speaking. This is to encourage learning the syntax of these vocabularies.

It is very important for the teacher to notice the three dimensions of using lexicon while assessing writing. A number of researchers have not found correlation between holistic assessment and lexical complexity (see Engber, 1995 and Wolfe-Quintero *et al.,* 1998), which, in my opinion, is something of significance that teachers, raters and assessors should attend to.

Finally, it is worth mentioning that the results of this study might help in teaching language in a way that could increase both accuracy and complexity in writing in the 'English as a foreign language' (EFL) context. They are also of use in writing assessment, leading assessors or raters to think more objectively and take different aspects of language into consideration while marking students' essays.

10.4 Future research

A number of important points that lie outside the scope of this research could be taken up in future research, as they are directly related to the possible applications of this research. The first one is an attempt to find out which one (or more) of the methods I have used most closely matches holistic scoring. This could be achieved by having a group of teachers score a number of the essays that are analyzed with these methods and then calculating the correlation between scores i.e. matching the holistic scoring with the analytic one. Taking this step will show how many of the features considered in the assessment of accuracy and complexity in this thesis have also been considered by the teachers while assessing essays holistically. This does not mean that only these features should be considered or they are the only correct aspects that teachers should bear in mind while assessing the students' essays. However, as the results of this research show features such as looking for correct sequences rather than only concentrating on errors and considering lexical and grammatical complexity as well as accuracy are important to consider while assessing students' written work holistically.

The second important aspect for future research is to try to automate the methods used in this thesis. If this is achieved, it will have far-reaching implications not only for the institutions where I collected my data but other institutions as well. This will save the time of the EFL teachers, allowing them to spend time on other areas of writing pedagogy. It will also help in bringing a more objective way of writing assessment than holistic rating. Nevertheless, even if this automated assessment were to be developed, it has to be treated with caution because automated assessment might pick up on some points and when students get these aspects right, they will get high scores even if they have not written the essays very well in terms of other aspects. Hence, teachers do tend to have a rather negative reaction to the notion of automatic essay assessment. As an EFL teacher, I support the view that automated assessment of essay writing has both advantages and disadvantages. However, the more developed and comprehensive the automated software is, the more reliable the automated assessment will be

and once a good system for assessment existed, it will really save a lot of time that the teachers could spare for other aspects of teaching.

The third point for possible further research is researching the way these methods can be applied in teaching writing and also speaking. This is an important future plan that I am interested to put it into practice. For this piece of future research, I am planning to investigate whether breaking down the students' essays into small units and giving them back to them to unite them into bigger units will result in any improvement of the accuracy and complexity of their writing. This could be done through having two groups; one control group that can be taught with any other method of teaching writing with the aim to increase accuracy and complexity and an experimental group that can be taught with this method. Then, the two groups could be compared for the improvement of accuracy and complexity in their writing. Other teachers could also be involved to test their reaction to this method of teaching writing. If the method worked well, it could be proposed on a larger scale like on the level of curriculum design and developing teaching materials.

The fourth point for further research outside the scope of this research is to apply the method of IAA in the assessment of the writing of lower level students i.e. students in their primary and secondary school stages. This will inform researchers on the levels of writing that could be analyzed with error analysis and the levels that could only be analyzed with IAA.

10.5 Conclusion

The starting point of this thesis was to assess accuracy and complexity in a set of essays written by 308 Kurdish university students in the context with which I am familiar. It presented particular challenges because the writing did not lend itself to the standard measures of normal error analysis, and I was therefore obliged by the data to devise novel ways of assessment.

This thesis has made a number of important contributions to the field of writing assessment and writing pedagogy. It has explored various methods and measures for assessing accuracy and complexity in English as a foreign language. The most outstanding contribution is looking at language from a positive perspective (correctness) instead of the negative perspective (error), hence the introduction of 'correctness analysis' as a counter-method to the 'error analysis'. The main point addressed in the thesis is the use of small units as units of

analysis in the method of 'correctness analysis' instead of T-units, clauses and sentences. Still related to the same point, the thesis has also provided the following mathematical formula for measuring the ratio of correctly structured words to the total number of words in a given essay:

$$accuracy = \frac{[no.of\ correct\ sequences\ (CS) \times (mean\ length\ of\ correct\ sequences\ (No.of\ words\ in\ CS/NCS)])}{total\ number\ of\ words\ in\ essay}$$

This formula measures accuracy giving a result between 0 and 1 with 0 meaning no word has been structured correctly and 1 meaning all the words are structured correctly i.e. slotted correctly in units of two words and more. The third main achievement of this study is the introduction and devising of An Integrated Approach to Achievement (IAA), a novel method for measuring accuracy in essay writing especially in low-level writing, which integrates multiple methods in only one method. The IAA involves both error analysis and correctness analysis in one method by taking error type and number into consideration and dividing the unanalyzed T units into small correct and 'almost' correct units. I am ambitious that this method may be turned in to an automatic tool for assessing the accuracy in writing and if this is achieved, it will have far reaching implications for language assessment. In regard to measuring complexity, this thesis has added to the area of writing assessment. The study has measured post modification by counting the ratio of different post nominal modifiers to the total number of noun phrases. The thesis has also added to the field of lexical complexity measurement by using a newly developed tool (the Lexical Complexity Analyzer) in measuring the lexical complexity of a number of essays. In writing pedagogy, this thesis has given over a considerable part of this chapter to the use of these methods in writing pedagogy. For example, teachers can make use of the small units produced by students to get them start putting these units in bigger units and hence help them learn the syntax of producing sentences.

This chapter has summarized the main conclusions of the present study. It has also shed light on the main pedagogical implications of this research for the teaching and assessment of EFL writing. One of the main implications for assessing writing is that it is more encouraging that students get positive feedback rather than negative feedback i.e. be informed where they have produced correct sequences rather than only highlighting errors to them. For writing pedagogy, it is significant that a sample of students' essays be divided into small units by the teacher and given back to the students for identifying the correct ones and correcting the incorrect ones. Then the students might be asked to combine these units into larger ones until

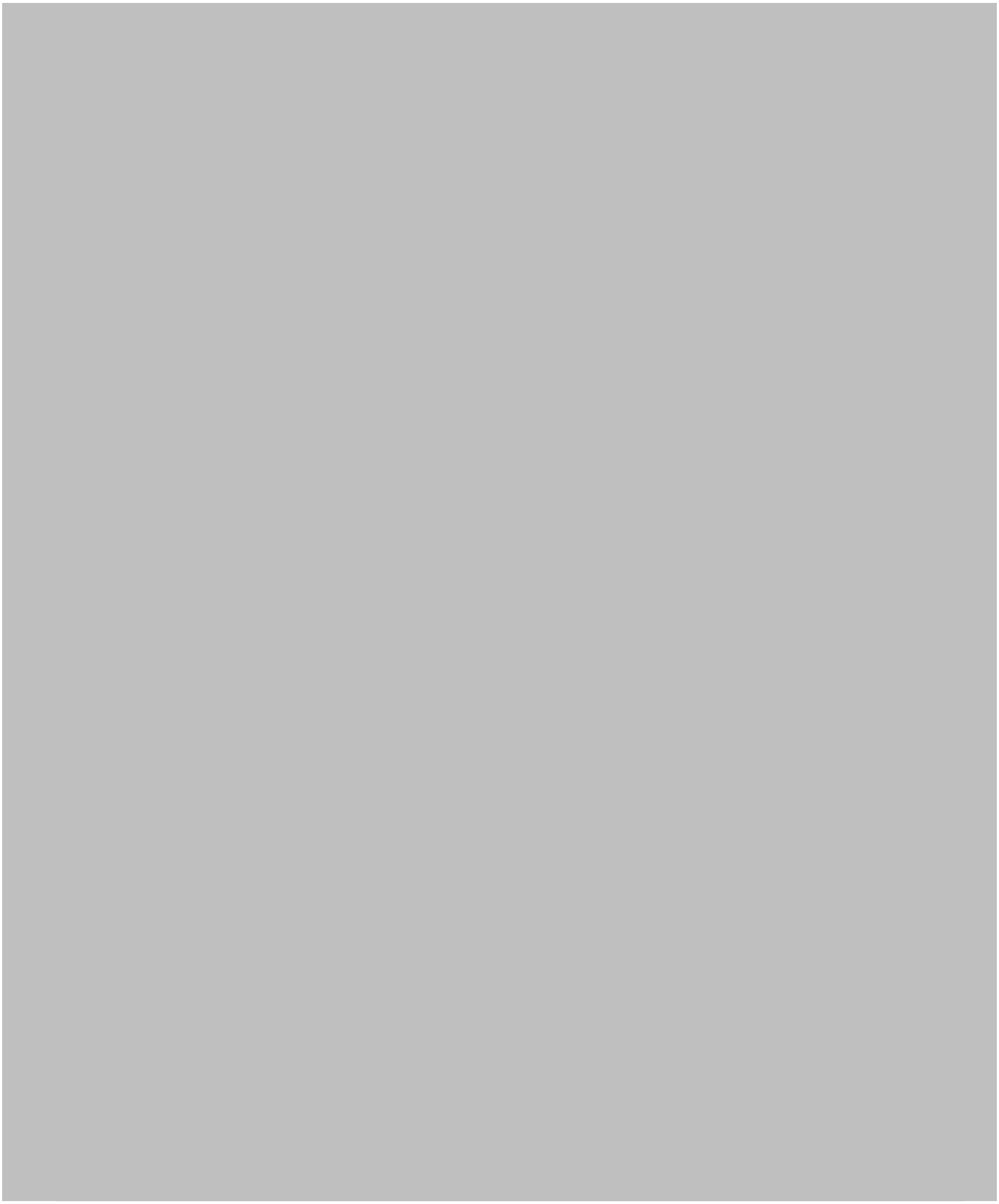
they end up writing a whole paragraph. Another conclusion that might be of a great benefit is that students be taught argumentative essays for learning how to write more complex sentences and learning more vocabulary as well being taught narrative essays for learning the use of verbs.

Notes

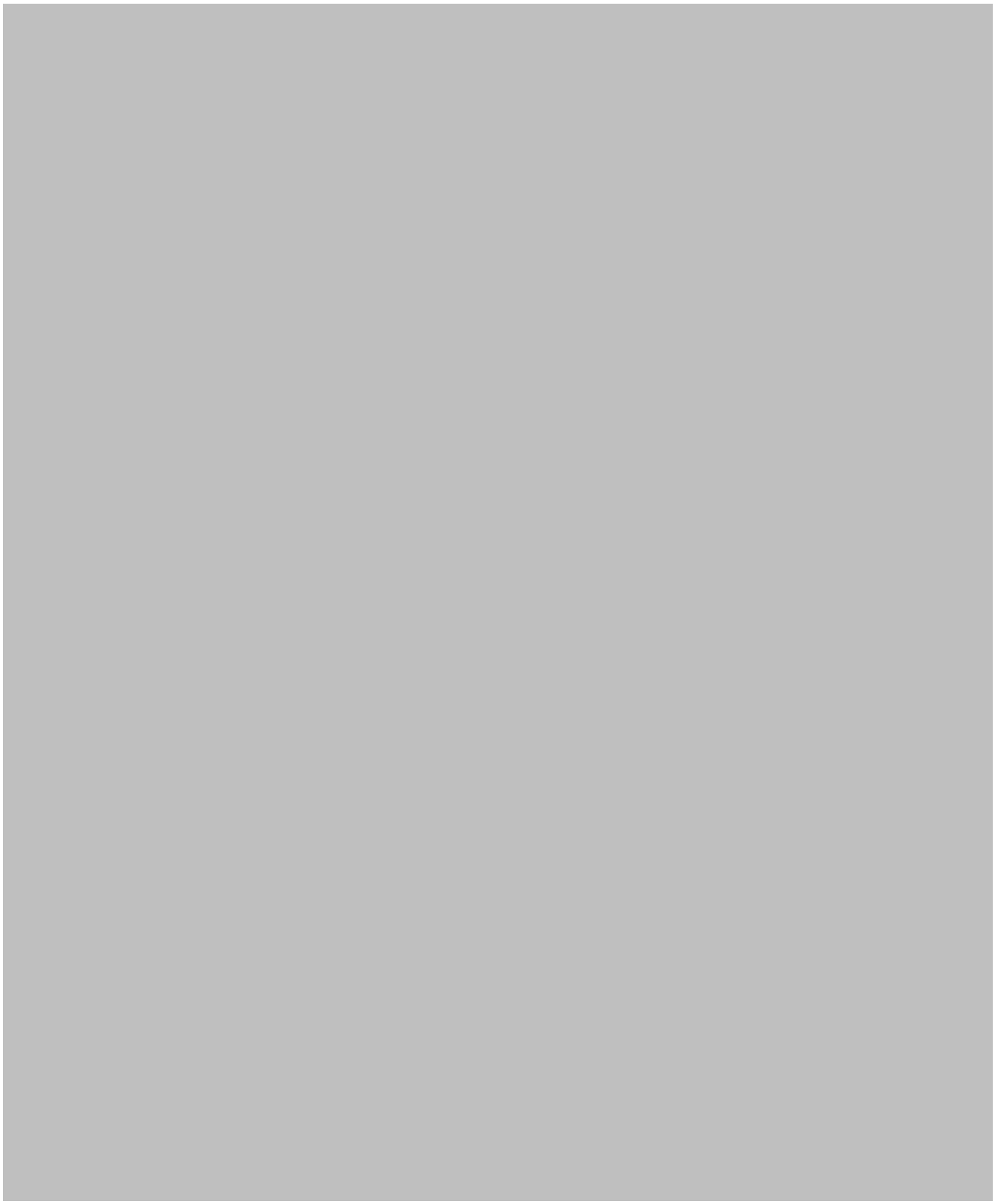
1. It can be calculated by dividing the total of error-free T-units by the total number of all T-units.
2. It can be calculated by dividing the total of error-free clauses by the total number of all clauses.

Appendices

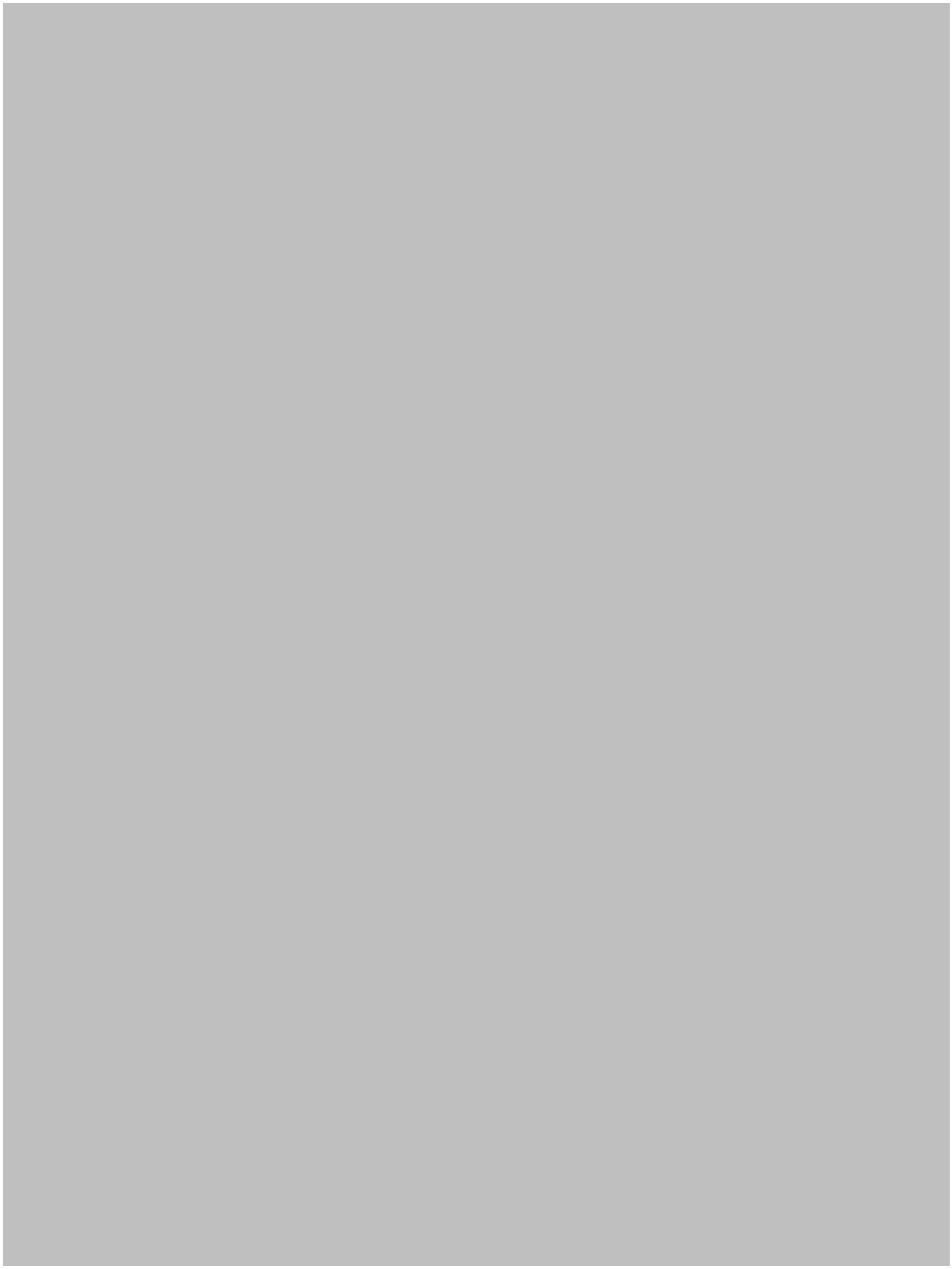
Appendix A



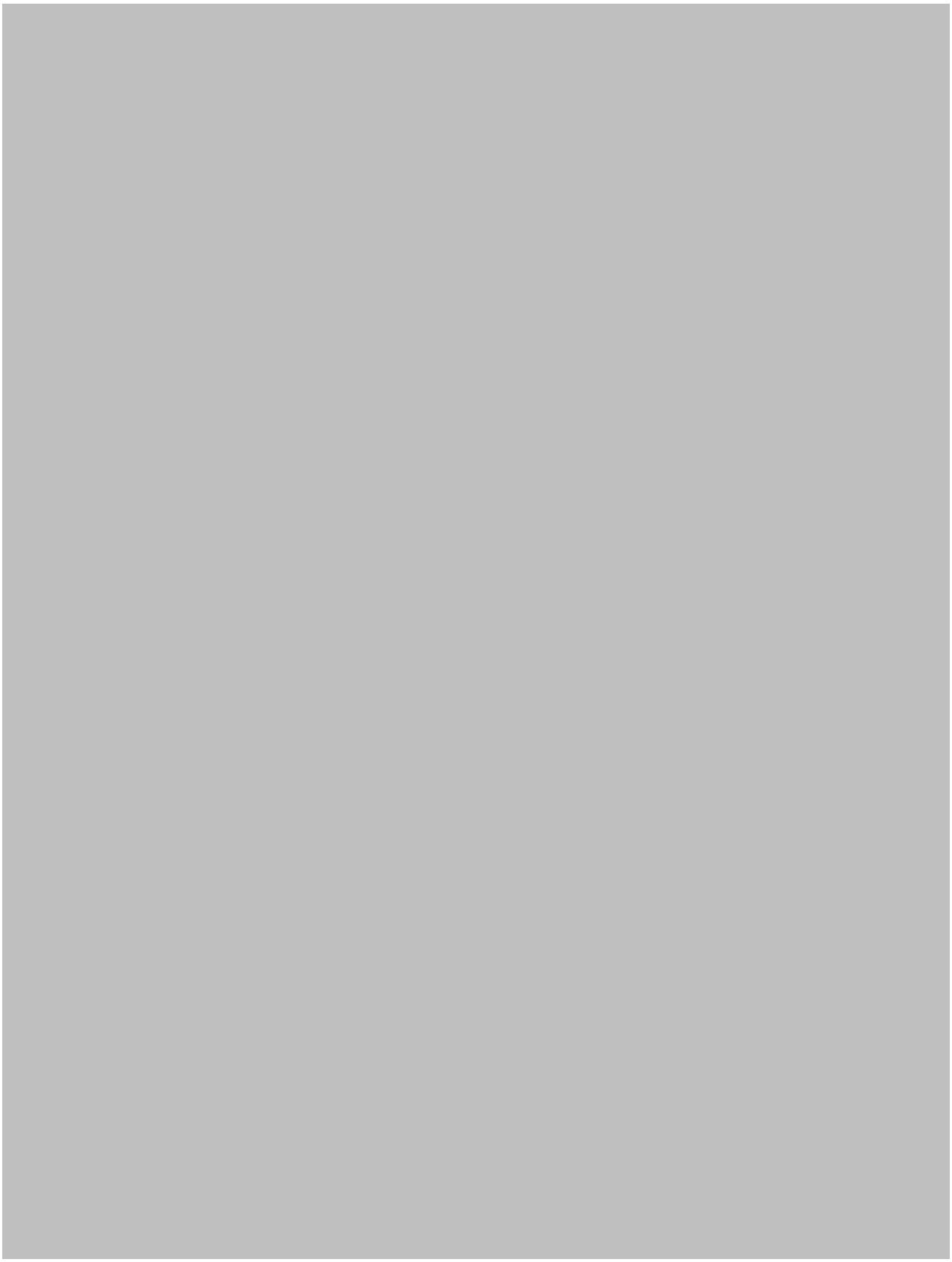
Appendix B



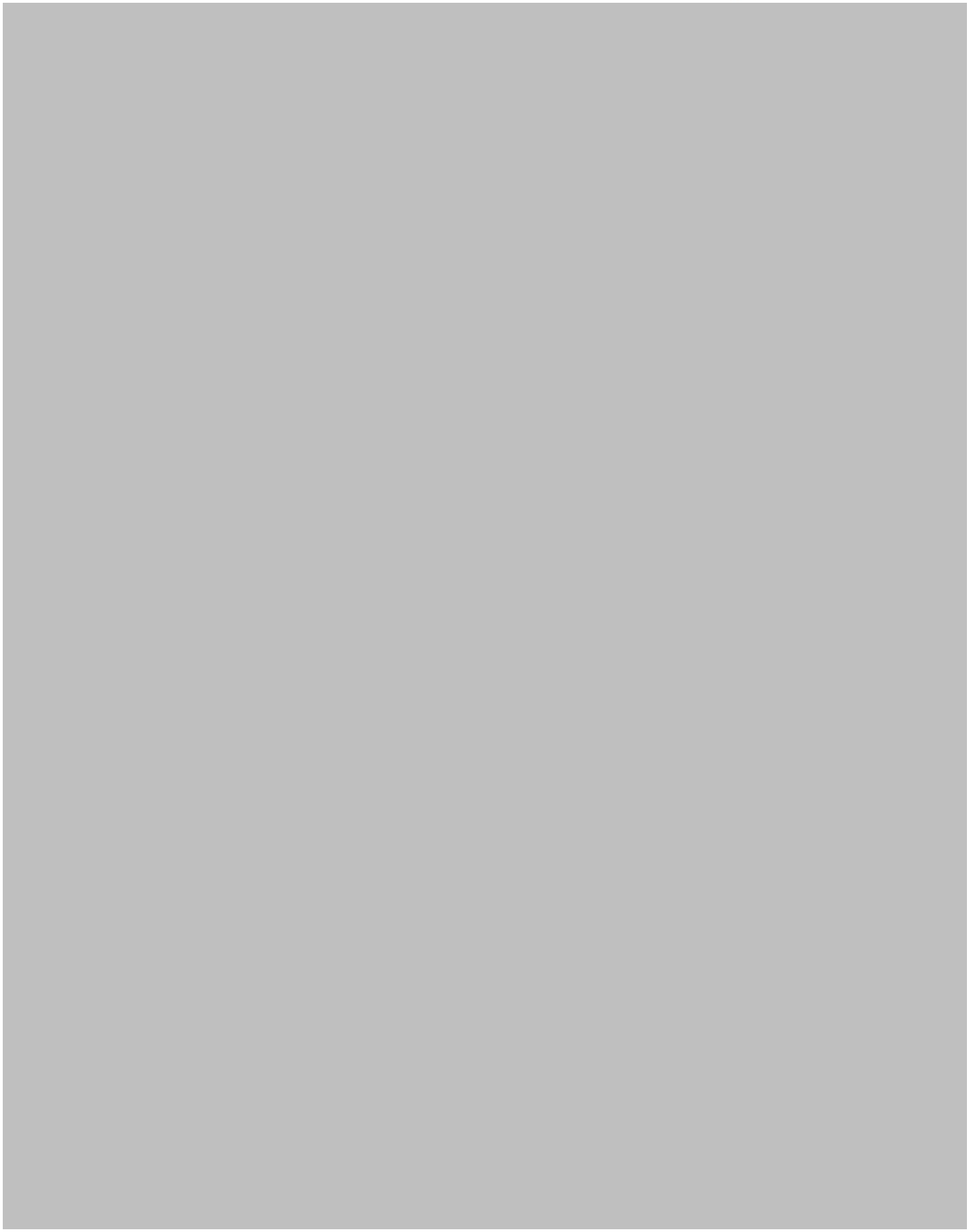
Appendix C



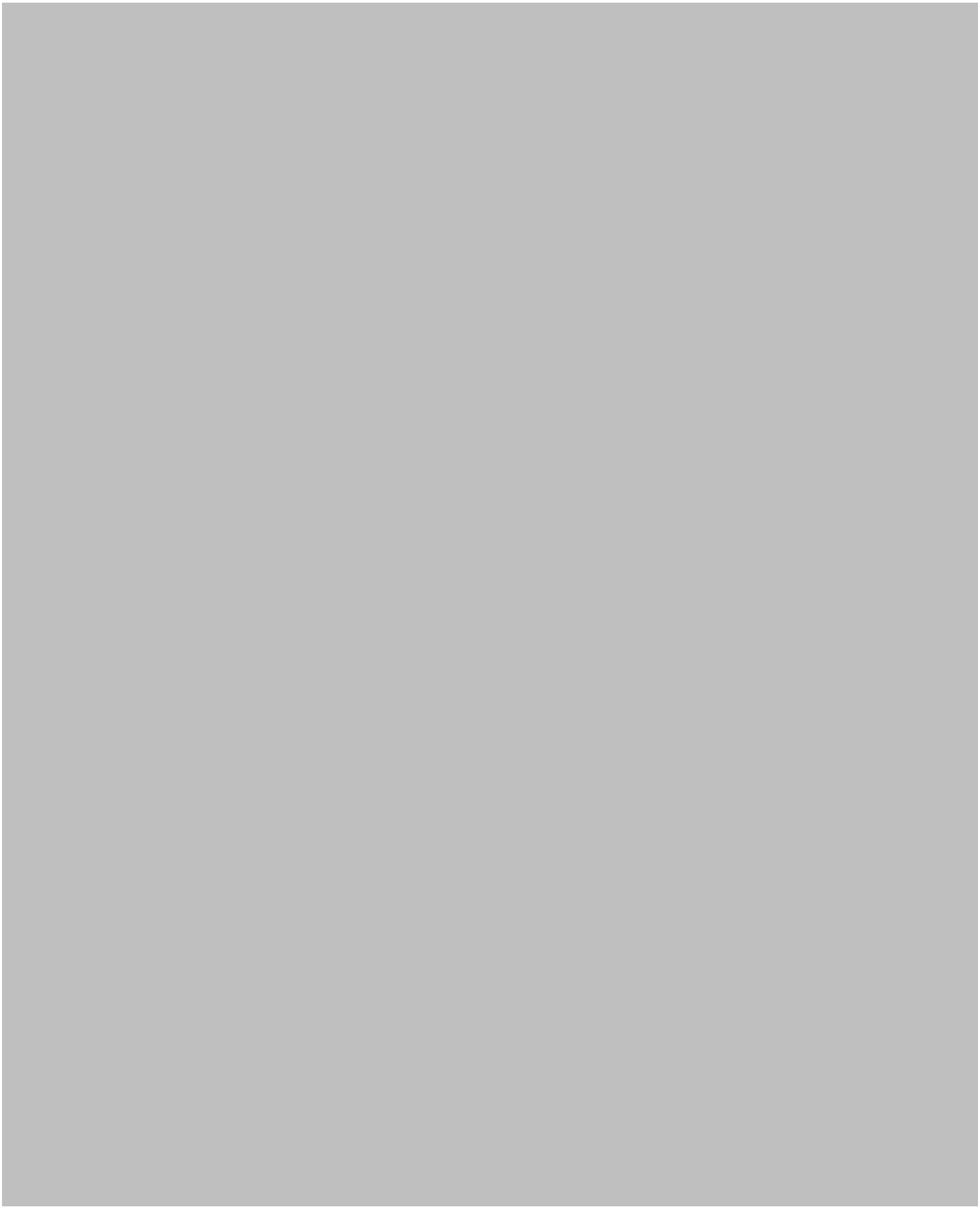
Appendix D



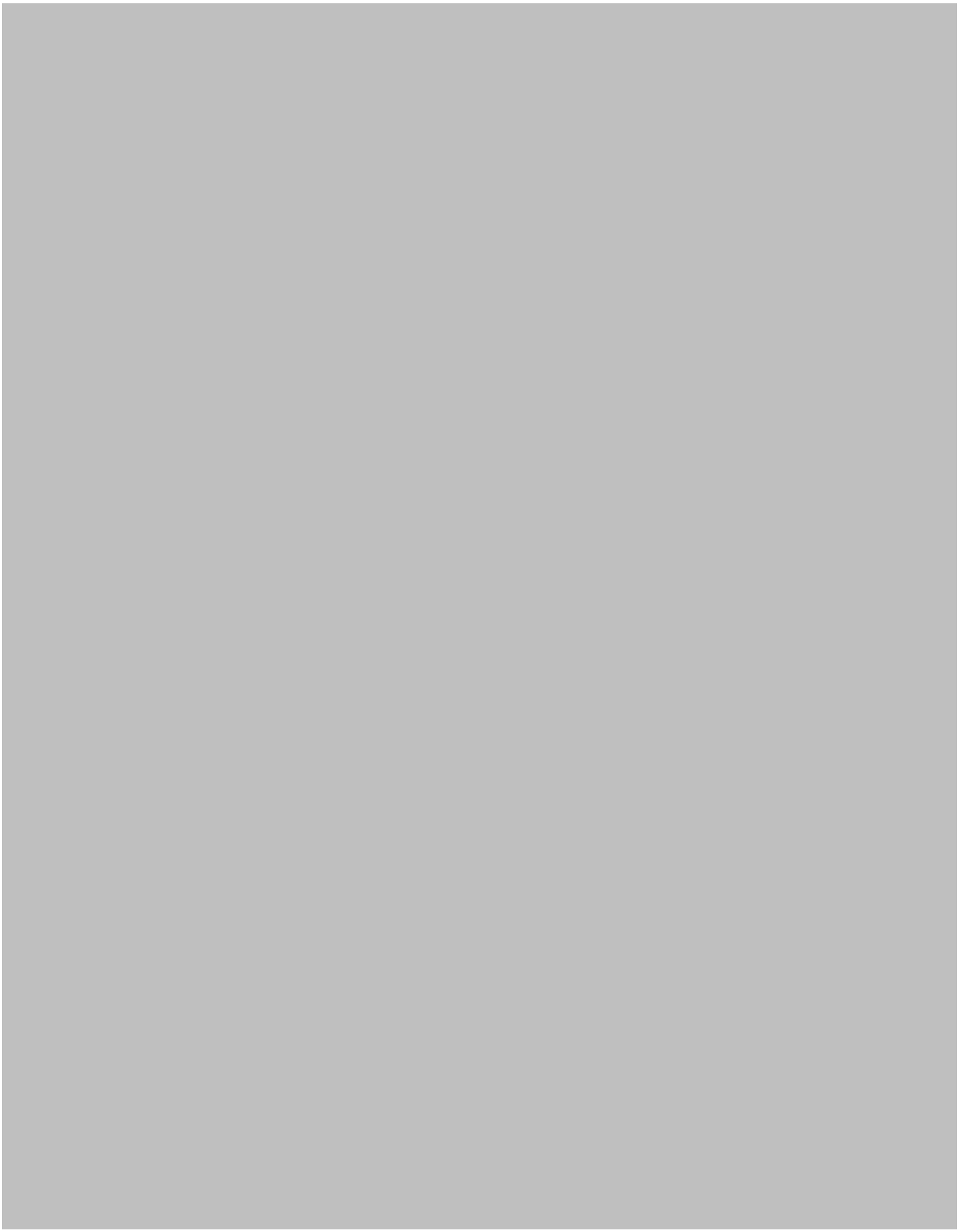
Appendix D

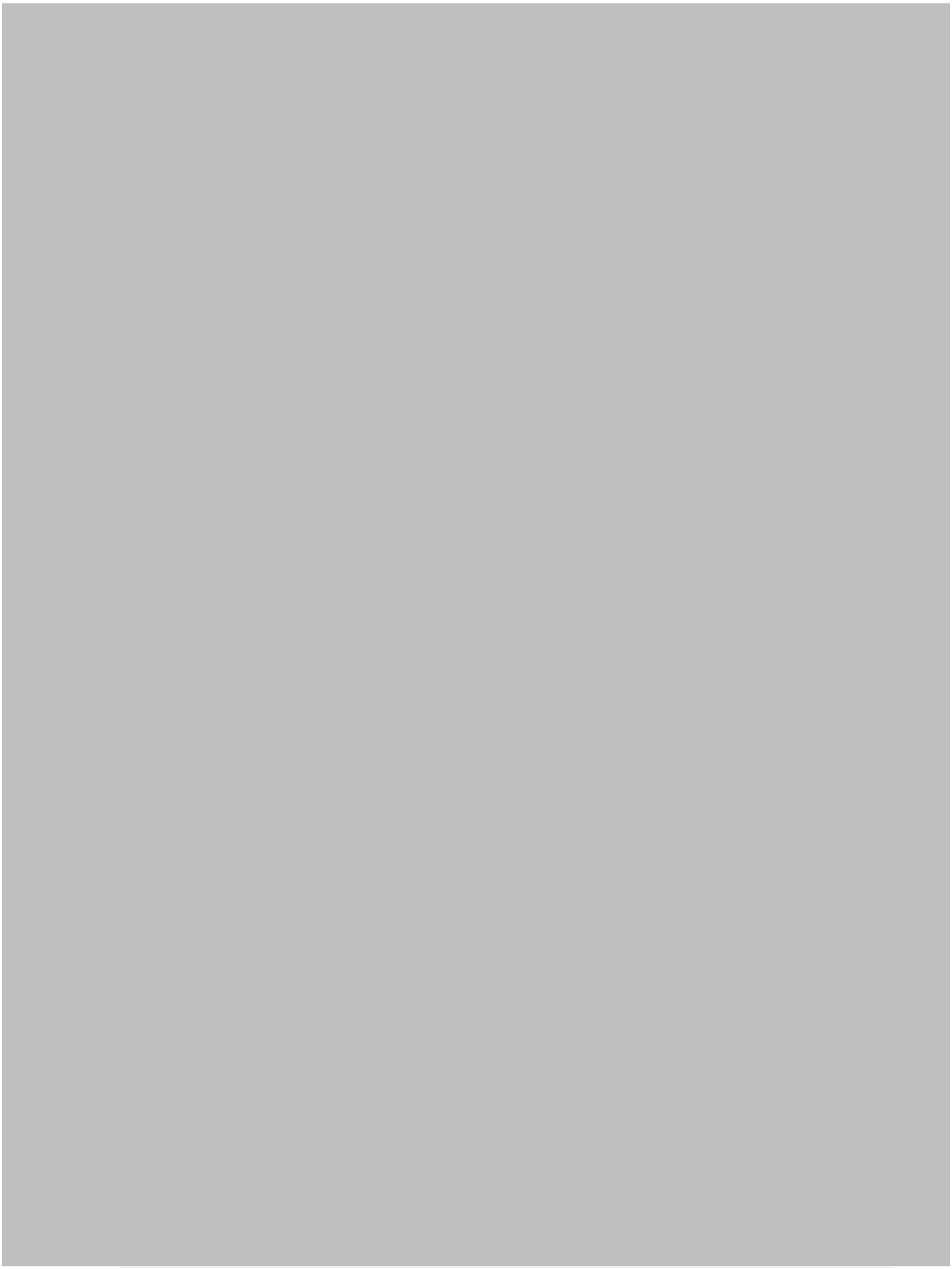


Appendix D

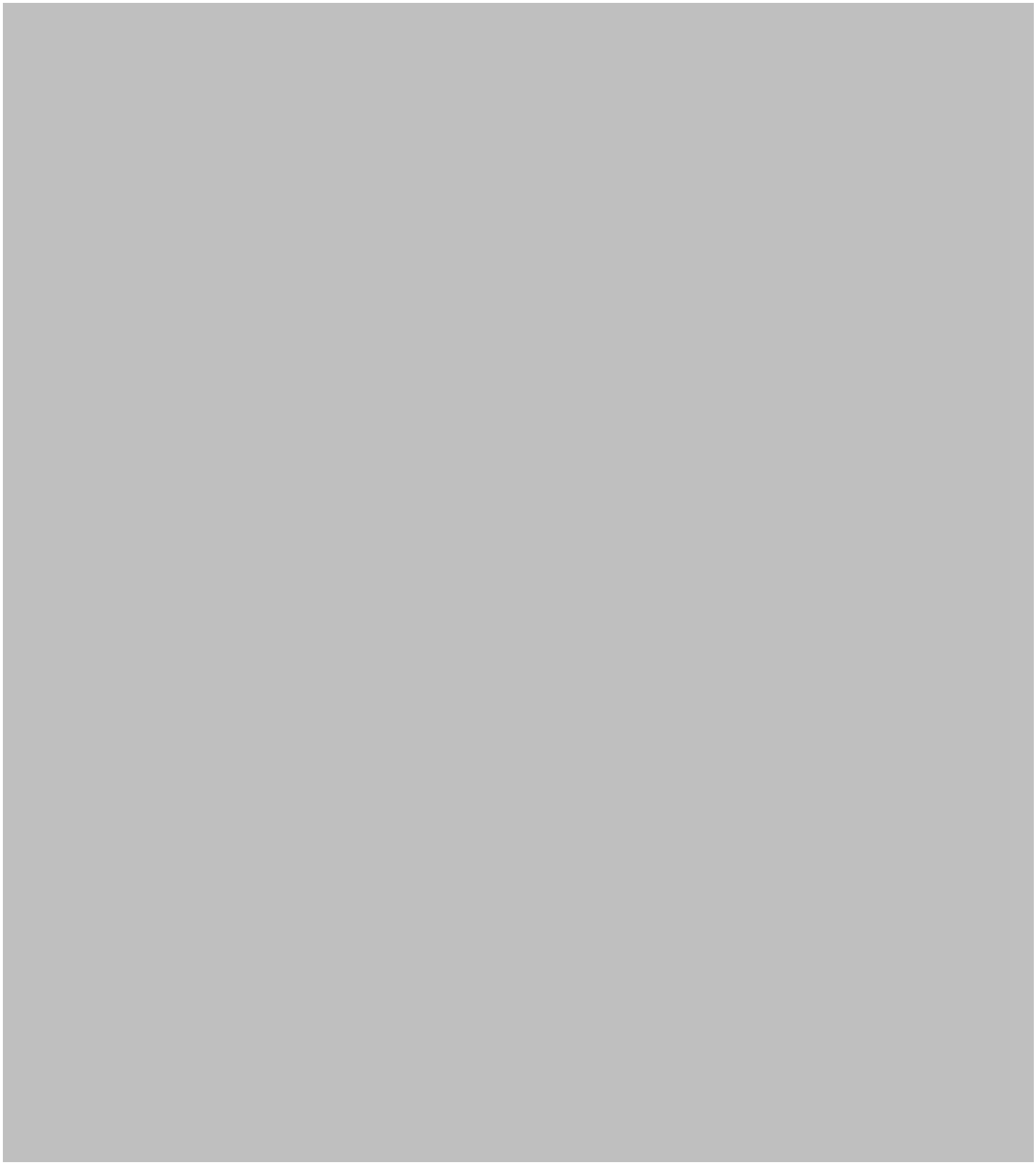


Appendix D



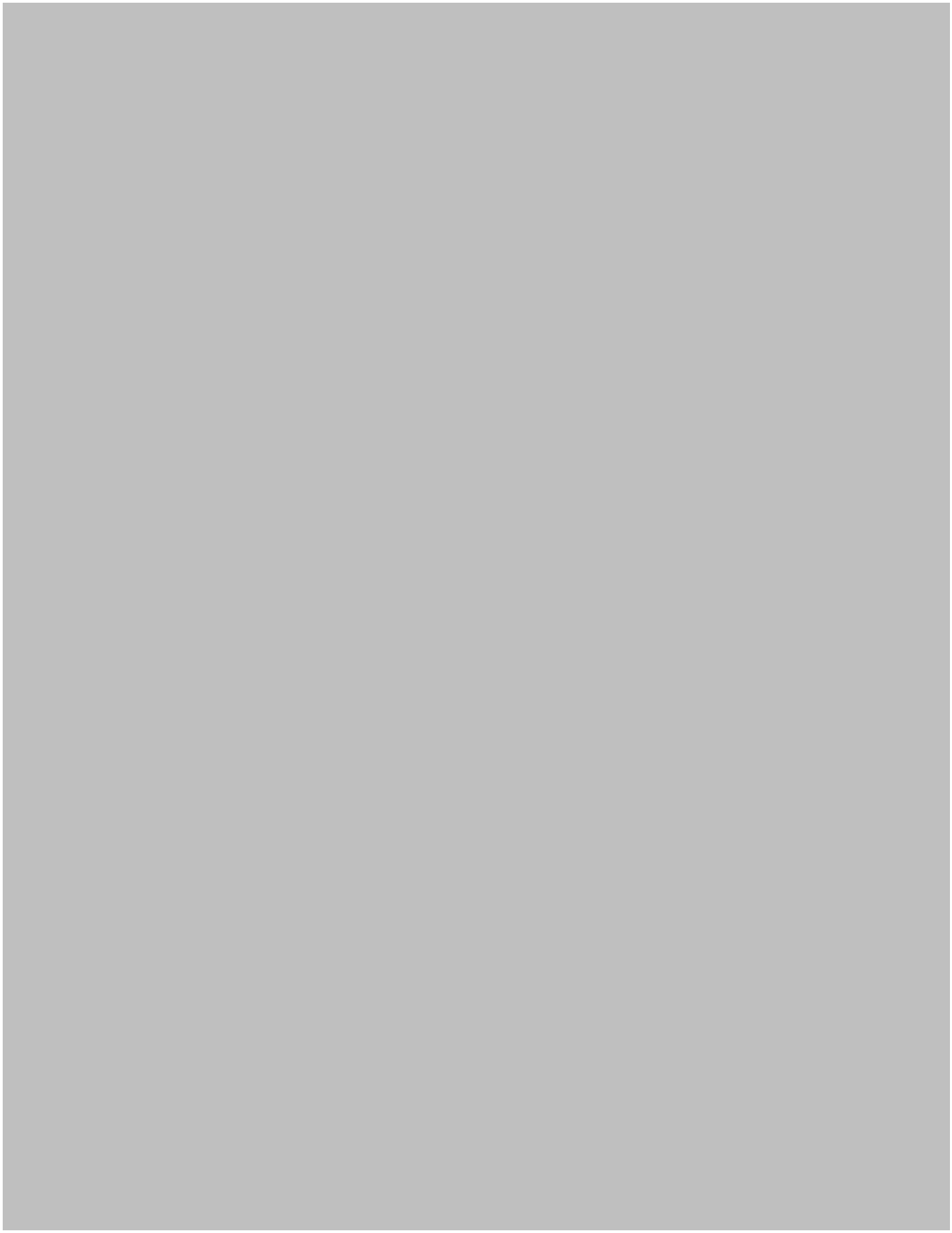


Appendix D

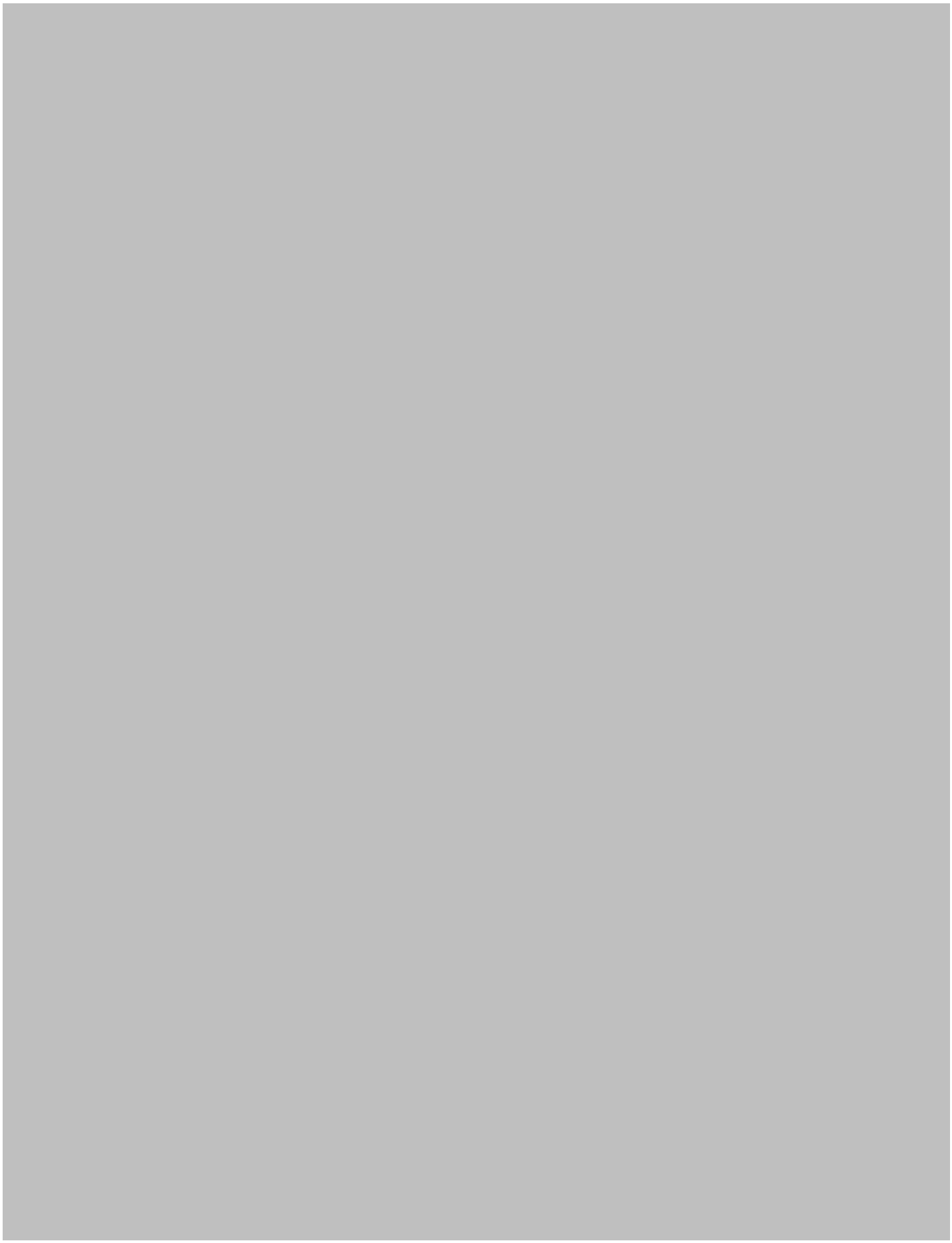


7

Appendix D



Appendix D



Appendix E

Dear Student,

I am a PhD student at the University of Birmingham in the UK. In my PhD I shall be investigating the writing of Kurdish-speaking learners of English. So, this PhD project needs your participation, which takes the form of essay writing. If you agree to participate in this project you are kindly requested to write one 400-500 word essay within a time limit of two hours on a topic of your choice from the attached lists (list A if you are in group A or B if you are in a group B).

Please also read the following important information:

- I will use your essay in my research, for my PhD thesis and for any publications arising from it.
- Your essay will not receive a mark, and taking part in this research will not count towards the assessment of your studies. If you choose not to write an essay, this will also not affect your marks for this course.
- Your essay will remain confidential to me. I may quote from it, but I will not give your name or any other details that might allow others to identify you.
- If you agree to take part and then change your mind, you can tell me within one month and I will remove your essay.
- If you wish, I will send you a link to any publications that appear as a result of this research.

If you agree to participate in this project, please confirm that you have read and understood the information above by signing your name below.

Full Name..... ..... Signature..... .....

Date.....

You can contact me using this information:

Haveen Muhamad Abdulmajeed
Phone no. (Kurdistan):
Phone no. (UK):
E mail:
Address:
52 Pritchatts Road
University of Birmingham
Edgbaston
Birmingham
B15 2TT

And contact my supervisor using the following information

Professor Susan Hunston
Telephone:
Email:
Address:
52 Pritchatts Road
University of Birmingham
Edgbaston
Birmingham
B15 2TT
UK

Appendix F

Full Name ……..

Age……

Gender……..

Mother tongue/dialect…….

Other foreign languages ……..

University …….. School…….. Year……

Years spent in an English speaking county……..

Appendix G

Error tagging workshop

Learner corpus summer school 2014

Sylviane Granger & Jennifer Thewissen

1. The Louvain error tagging guidelines

1. Error detection vs rewriting

Students' predisposition are not taken into account.
REWRITING: *Aptitude for teaching is not considered.*

2. Error tagging principles

Principle 1

Do not tag on the basis of the corrected/targeted word/ phrase, but on the basis of the incorrect word/ phrase only.

e.g. *The main feature of a campus like Louvain-la-Neuve is the conviviality.*
Correction: *The main feature of a campus like Louvain-la-Neuve is its conviviality.*

This error should be tagged (GA) (misuse of article) rather than (GDO) (misuse of a possessive determiner):
The main feature of a campus like Louvain-la-Neuve is (GA) the its conviviality.

Principle 2

Place the tag immediately before the error (word or phrase) that needs to be corrected.

e.g. *This type of exchanges.*
Correction: *this type of exchange*
This type of (GNN) exchanges $exchange$

exception

The X* tags (lexico-grammatical errors), i.e. (XNPR), (XVPR), (XNCO), (XVCO), (XNUC) etc., should be placed in front of the word that triggers the error rather than in front of the error itself.

Appendix G

e.g. *Students have the* (XNCO) *possibility to leave $possibility of leaving$...*
and **not** *Students have the possibility (XNCO) to leave $possibility of leaving$...*

Principle 3

Place the correction immediately after the erroneous word/phrase. For retrieval purposes, the corrected form is preceded and followed by a $ sign.

e.g. *some* (XNUC) *advices $advice$.*

Principle 4

When there are two types of errors in the same word/phrase, double tag this word/phrase.

≤ sometimes several errors on same word

e.g. they *puts forward*
Correction: *they maintain*
There is
1) a problem of concord between subject and verb and
2) a problem of lexical choice (phrasal verb).

This error should be tagged:
they (GVN) puts forward $put forward$ (LP) puts forward $ maintain$

Principle 5

Use the 'zero' (0) to indicate a missing word or a correction that consists of the deletion of a word /phrase.

e.g. *not only does* (GA) *0 the economy take over...*
in order to avoid (GP) *any 0 conflicts.*

Principle 6

Do not tag errors resulting from the correction of a previous word

e.g. (GNN) *Cigarette $Cigarettes$ is dangerous for your health.* ⚠ because the student didn't produce that error.
and **not** (GNN) *Cigarette $Cigarettes$* (GVN) *is are dangerous for your health.*

exception

Appendix G

Correcting a **tense error** may lead to the need for other tense changes even though in the actual context there is no error. In such cases, **all auxiliaries and tense forms that need changing should be tagged separately.**

e.g. *What* (GVAUX) *shall $would$ we think for example if we* (GVT) *see saw English people driving on the right ?*

3. Errors vs. infelicities

Version 1.3 of the tagging manual makes a crucially important distinction between **real errors**, on the one hand, and **infelicities**, on the other. Real errors involve **the breach of a specific linguistic rule** (grammatical, lexical or other), while infelicities refer to **non-erroneous, but odd-sounding language**. Infelicities include the following:

1. **Register problems** such as contractions (*couldn't, isn't*), informal expressions (*a bit, lots of,...*), geographical differences (*colour/color*), slang, etc.
2. Questions of **political correctness** (e.g. *'chairman'* instead of the politically correct *'chairperson'*, *'mankind'* instead of *'humankind'*, *he/she, himself/herself*, etc.).
3. **Stylistic problems** (e.g. *What else could be the reason for this resistance...* instead of *What other reason could there be for this resistance...*).

2. The Louvain error tagging manual 1.3.: 53 error tags

Error tags and tag name	Tag description	Example
	Formal errors	
FM: Form, morphology	Morphological error resulting in non-existent English word	*It is (FM) unpossible $impossible$*
FS :Form, spelling	Spelling error resulting in non-existent English word	*The fast spread of television can transform it into a double-edged (FS) wheapon $weapon$.*
	Grammatical errors	
GA: Grammar, article	Article errors involving the definite, indefinite or zero article (sentence grammar error)	*(GA) The 0 life is beautiful.*
GDD: Grammar, determiner, demonstrative	Errors on demonstrative determiners	*(GDD) This $These$ elements cannot be separated.*
GDO: Grammar, determiner, possessive	Errors on possessive determiners	*People accept jobs according to how much they get paid but not according to (GDO) his $their$ preferences.*
GDI: Grammar, determiner, indefinite	Errors on indefinite determiners	*He does not have (GDI) some any expectations.*
GDT: Grammar, determiner, other	Errors on determiners other than those listed above	*(GDT) Whichever $Which$ man was it you saw yesterday?*

Appendix G

Error tags and tag name	Tag description	Example
GADJCS: Grammar, adjective, comparative/superlative	Errors on the comparative or superlative use of an adjective	*The role that women should play in a (GADJCS) more fair $fairer$ present-day society*
GADJN: Grammar, adjective, number	Adjective erroneously used in the plural	*The last sentences have been (GADJN) favourables $favourable$ to women.*
GADJO: Grammar, adjective, order	Erroneously ordered adjectives	*A (GADJO) leather black small $small black leather$ handbag*
GADVO: Grammar, adverb, order	Misplaced adverb	*They (GADVO) see only $only see$ other criminals.*
GNC: Grammar, noun, case	Errors involving the use of the Saxon genitive	*Behind the (GNC) Berlin's wall $Berlin wall$*
GNN: Grammar, noun, number	Addition or omission of the plural morpheme on nouns	*Bearing in mind that sex equality is one of the great (GNN) reason $reasons$ for fights in most places around the world (...)*
GPD: Grammar, pronoun, demonstrative	Errors on demonstrative pronouns	*What is harassment? The dictionary says (GPD) that it is a behaviour which is intended to trouble or annoy someone.*
GPP: Grammar, pronoun, personal	Errors on personal pronouns	*The big majority of children have a computer or a video-game, with which (GPP) 0 $they$ spend (waste, in my opinion) a great number of hours.*
GPO: Grammar, pronoun, possessive	Errors on possessive pronouns	*My computer did not cost as much as my sister's. (GPO) His $Hers$ was more sophisticated.*
GPI: Grammar, pronoun, indefinite	Errors on indefinite pronouns	*A lower-class man is not on an equal footing with his middle- or upper-class (GPI) one $counterpart$.*
GPF: Grammar, pronoun, reflexive/reciprocal	Errors on reflexive or reciprocal pronouns	*They didn't need to communicate (GPF) themselves 0 outside their homes.*
GPR: Grammar, pronoun, relative/ interrogative	Errors on relative or interrogative pronouns	*The government took several measures to stop the strikes, (GPR) that $which$ was not effective.*
GPU: Grammar, pronoun, unclear reference	Use of a pronoun whose reference is unclear	*But there are also imprisoned people waiting for their execution who are innocent. They never had a fair trial and a real chance to get out of (GPU) it $jail$. These people often do not have enough money to get their own attorney.*
GVAUX: Grammar, verb, auxiliary	Misuse of modal auxiliaries (*can, should, may*,etc.), primary auxiliaries (*be, do, have*) or semi-auxiliaries (*ought to, used to, dare, need*)	*So if there is an army it (GVAUX) might $should$ be professional, and formed by people who believe in that and want to dedicate their lives to it.*
GVM: Grammar, verb, morphology	Erroneous use of existing verb forms (e.g. a simple past form instead of a past participle form, an infinitive instead of a past participle)	*It is generally (GVM) agree $agreed$ today that we live in a world where television plays an important part.*
GVN: Grammar, verb, number	Subject-verb agreement errors	*How do you think that a man (GVN) react $reacts$ when he hears*

4

Appendix G

Error tags and tag name	Tag description	Example
		that a woman is shocked when she receives a letter beginning with "Dear Sirs"?
GVNF: Grammar, verb, non-finite/finite	Errors involving non-finite/finite verb forms	*(GVNF) To travel $Travelling$ by public transport is recommended.*
GVT: Grammar, verb, tense	Erroneously chosen tense or aspect	*He learned a profession in the prisson, and now he (GVT) wrote $writes$ poetry and (GVT) took $takes$ part in the publication of a prisson's journal.*
GVV: Grammar, verb, voice	Use of the active instead of passive voice or passive instead of active voice	*This seems impossible to (GVV) be achieved $achieve$.*
GWC: Grammar, word class	Inappropriate use of a word class: adjective used instead of a noun, adverb instead of an adjective, etc.	*We are going to review the following subjects: Labour discrimination, the right to vote, the fight against male (GWC) chauvinist $chauvinistic$ behaviours.*
	Lexical errors	
LCC: Lexis, conjunction, coordination	Errors on coordinating conjunctions (*and, but, neither..nor, etc.*)	*Life is not only work (LCC) or and study.*
LCLC: Lexis, connector, logical, complex	Errors on multi-word logical connectors	*(LCLC) In consequence to this $Consequently$, people are very busy and tired.*
LCLS: Lexis, connector, logical, single	Errors on single-word logical connectors	*It is not a very nice flat. (LCLS) Moreover $However$, it is cheap.*
LCS: Lexis, conjunction, subordination	Errors on subordinating conjunctions (*if, whether, that, etc.*)	*We discussed (LCS) if $whether$ we should close the shop.*
LP: Lexical phrase errors	Errors on multi-word units (lexical bundles, phrasal verbs, etc.)	*I was riding my bike through the village when I met a Turk, who was (LP) of middle age $middle-aged$.*
LS: Lexical single errors	Conceptual, collocational or connotative lexical errors on single existing English words	*Resorting to violence might be somehow a (LS) comprehensive $understandable$ reaction.*
	Punctuation errors	
QC: Punctuation, confusion	Confusion between two punctuation markers (e.g. run-on sentences where a comma is used instead of a full stop)	*Some creative people make the most of their spare time by imagining or even building things (QC), $.$ others create works of art which are fruits of their inmost beings.*
QL: Punctuation, lexical	Punctuation marker used instead of lexical item (usually a conjunction of coordination) or lexical item instead of punctuation marker	*He took the books (QL) and $,$ the records and the computers.*
QM: Punctuation, missing	Missing punctuation marker	*Physically (QM) 0 $,$ you do not run any risks but it is very dangerous for your mind.*
QR: Punctuation, redundant	Redundant punctuation marker	*Women were seen conducting affairs, and bringing negotiations to satisfactory conclusions in what men always claimed to be (QR) :*

Appendix G

Error tags and tag name	Tag description	Example
		0 "a man's world".
	Style/sentence errors	
SI: Sentence, incomplete	Fragments such as verbless sentences	*(SI) Another example. $Another example is:$ Yesterday we spoke about the Gulf War (…)*
SU: Sentence, unclear	Incomprehensible sentence	*(SU) Beggars of reflection have power and very often they use it in such a wrong way that make of imagination become slaves or just disappear $?$.*
	Word missing/redundant/order errors	
WM: Word missing	Omission of words, except pronouns, dependent prepositions, articles, connectors, auxiliaries	*The future soldiers make an strict physical training and (WM) 0 sit some exams.*
WO: Word order	Problems of word order that do not fall into the categories of Adverb Order (GADVO) or Adjective Order (GADJO)	*Think about (WO) how would be your house $how your house would be$ without the last century's inventions.*
WRS: Word redundant singular	Unnecessary use of a single word, except articles (GA), connectors (LC*), pronouns (GP*), dependent prepositions (X*PR), and auxiliaries (GVAUX)	*Actual life is very complicate and (WRS) extremely 0 full of worries.*
WRM: Word redundant multiple	Unnecessary use of multiple words, except articles (GA), connectors (LC*), pronouns (GP*), dependent prepositions (X*PR), and auxiliaries (GVAUX)	*Others comment that (WRM) the fact is that 0 once you are inside, if you like it, you can even re-enlist.*
	Lexico-grammatical errors	
XADJCO: Lexico-grammar, adjective, complementation	Erroneous complementation of adjectives	*Both sex (male and female) are (XADJCO) capable to do $capable of doing$ military service.*
XADJPR: Lexico-grammar, adjective, preposition	Adjective used with an erroneous dependent preposition	*How many public places are easily (XADJPR) accessible for $accessible to$ wheelchairs?*
XCONJCO: Lexico-grammar, conjunction, complementation	Erroneous complementation of conjunctions	*I keep thinking that television causes violence (XCONJCO) rather than discouraging $rather than discourage$ it.*
XNCO: Lexico-grammar, noun, complementation	Erroneous complementation of nouns	*Students have the (XNCO) possibility to leave $possibility of leaving$.*
XNPR: Lexico-grammar, noun, preposition	Nouns used with an erroneous dependent preposition	*He has a (XNPR) thirst of $thirst for$ knowledge.*
XNUC: Lexico-grammar, nouns, uncountable/ countable	Errors involving the countable use of uncountable nouns	*The tremendous (XNUC) progresses $progress$ realized by science have disrupted our habits and our way of living.*
XPRCO:	Erroneous complementation of prepositions	*The result is an invention that (XPRCO) instead of help $instead of*

6

Appendix G

Error tags and tag name	Tag description	Example
Lexico-grammar, preposition, complementation		*helping$ society, damage it.*
XVCO: Lexico-grammar, verb, complementation	Erroneous complementation of verbs	*What about the people who cannot (XVCO) afford going $afford to go$ to these kind of centres?*
XVPR: Lexico-grammar, verb, preposition	Verb used with an erroneous dependent preposition	*The classroom must often have (XVPR) resembled to $resembled$ a "Chamber of Horrors".*
(Infelicities)		
Z Infelicities	Not full-blown errors, but rather 'milder' problems involving register, questions of political correctness and stylistic problems (e.g. long-winded introductory sentences)	*(Z) kids $children$* *(Z) he $he or she$*

GVT => ing + ed tense and aspect error

7

Appendix H

Essay no.	Codes												
	FS	FM	GA	GDD	GDO	GDI	GDT	GADJCS	GADJN	GADJO	GADVO	GNC	GNN
{1}	4	1	1					1			1	2	5
{2)			2										1
{3}	8		1	1			1	3					1
{4}	2		9		3	1						1	2
{5}	5		4	8	1								7
{6}	1		8	1									2
{7}	3	1	3						1				3
{8}	21	1	6			1						1	3
{9}	5		3										3
{10}	4		7										
{11}	2		6					5	1			1	7
{12}	1		3						1			1	6
{13)	3												
{14}	13	4	2	1				1					
{15}	2		1										1
{16}	5		7										4
{17}	4		12										1
{14}	13	4	2	1				1					
{15}	2		1										1
{16}	5		7										4
{17}	4		12										1
{18}	10		4		1								1
{19}	5	2	2	1		1							9
{20}	6												
{21}	6	1	3										
{23}	8	4	8										4
{24}	5		3										
{25}	12		3	3									2
{26}	7		1										1
{27}	1	2											1
{28}	7		5		1							1	2
{29}	4		3			1			1	1		2	2
{30}	3			1									3
{31}	4	1	3		1								9
{33}	11		6		2			1					
{34}	1		4										6
{35}		2	5								1	1	5
{36}	5	1	11			1		2					2
{37}			8	3									
{38}	1		9	1									5
{39}	7		3					1	6			2	2
{40}	1		6										1

1

Appendix H

Essay no.	Codes												
	FS	FM	GA	GDD	GDO	GDI	GDT	GADJCS	GADJN	GADJO	GADVO	GNC	GNN
{41}	10		12										2
{42}	20	1	7									2	4
{43}	3		2										4
{44}	13		18										22
{45}	8		6	5				1				1	5
{46}	5	3	5					2					7
{47}	15		23					1					1
{48}			14			1			2				6
{49}	3	1	3										3
{50}	3		3										2

2

Essay no.	Codes												
	GPD	GPP	GPO	GPI	GPF	GPR/I	GPU	GPR	GVAUX	GVM	GVN	GVNF	GVA
{1}								3		1	1		8
{2}	2	1						1		3			3
{3}		11		1				5		5	3		3
{4}		6						5	3		1		
{5}		1				1			4	7			1
{6}		1		1						2			
{7}		2							1	3	1		
{8}		2					1	2		1	3		
{9}				1			1		1				
{10}	1	3					1	2	1		1		
{11}						1				1	3		
{12}		2			1				1	1	5		
{13)								1					
{14}		1				1							3
{15}		2				1		1	1				
{16}	1							1		1	5		1
{17}										1	1		
{14}		1				1							3
{15}		2				1		1	1				
{16}	1							1		1	5		1
{17}										1	1		
{18}		3				1		3	6	1	7		
{19}		6		1		1	3	3	2		3		
{20}		2											
{21}		3		1				1	1	4			1
{23}		6		1				4	4				1
{24}										1	1		
{25}		1		1						1	2		1
{26}		3		2		1	1	1		1			
{27}		1											
{28}						1							
{29}		1						3		1			
{30}									2		1		
{31}		2				2		3					1
{33}	1	5		1				4			5		
{34}								5			4		
{35}				1	1	1		1	1	3	1		
{36}							2	3	1		19		1
{37}		3		1				1	1	1	4		3
{38}						2		7	1	2			
{39}		5		1				5	3		6		
{40}		1					1	3			1		

Appendix H

Essay no.	Codes												
	GPD	GPP	GPO	GPI	GPF	GPR/I	GPU	GPR	GVAUX	GVM	GVN	GVNF	GVA
{41}		2						3	1	1	1		
{42}	1	3	2	5				5		2	1		1
{43}								4			4		
{44}		2						6	1		1		
{45}				1				6			6		
{46}				2			2	1			3		
{47}			1					2		1	3		3
{48}	1	2		1			1	1	2		1		
{49}								2			1		1
{50}								2					

Essay no.	Codes														
	GVT	GVV	GWC	LCC	LCLC	LCLS	LCS	LP	LS	QC	QL	QM	QR	SI	SU
{1}	2			1	1	2	1	2		1		5	3		
{2)	2			1		1	1	7	7	5		2	1		2
{3}	5	1	3				1	1	6	8			2	3	
{4}	10	1	2	1				2	9	6		6	4	2	
{5}	9		1	3				3	2	5		5	5	6	7
{6}	2	1	2	3	2	1	1	5	11	17		1	1	2	2
{7}	5		1	1		1		3	3	4		2		1	
{8}	1			1				5	2	4			2		
{9}	3		1					7		5		4	1	1	1
{10}		1						2	2	9		4	6	1	
{11}	7		1							4		7		1	
{12}	1		3	1				7	2			2		1	
{13)	1									2		1		1	
{14}	1						1	4	4	5		9		1	
{15}		1	2						1	8		3		2	2
{16}		1	1	2				4	2			17		2	4
{17}	3	2	1					2	2	3		12	1		
{14}	1						1	4	4	5		9		1	
{15}		1	2						1	8		3		2	2
{16}	1	1	1	2				4	2			17		2	4
{17}	3	2	1					2	2	3		12	1		
{18}	2	1	1					1	2	8			1		
{19}	3		3	1			1	4	1	1	1	8	2		
{20}				2		1		2		1				2	1
{21}	4			1				5	8	6		4	1		
{23}	4	6	8		1			2	10	21			11		
{24}	2		2	1	2			1	1		1	7			2
{25}	3							5	2	12		2	10	1	
{26}	6		2						3	10		5	2	2	
{27}	2		1	1					1			3			
{28}	1	1		3				6	4	1		1	3	2	
{29}			1	2		1		9	5	3		6	1		
{30}			4	2				7	3			1	1	2	2
{31}	2			3				7		3	2	20		1	
{33}			2	1	1			8				10			
{34}			1					1	1	1		2	1	2	1
{35}	1		5		1			5	1	3		4			2
{36}			2		1			1	7	4					
{37}	1							2	1			2	3	7	
{38}			2	1				1	2	2		8			
{39}	2		9				1	2	1			9			
{40}								3	7	4		15			1

Appendix H

Essay no.	Codes														
	GVT	GVV	GWC	LCC	LCLC	LCLS	LCS	LP	LS	QC	QL	QM	QR	SI	SU
{41}	1		1					1	2			4		1	
{42}	5	1	4									12			
{43}	1	1	2	1				3	2			8			7
{44}			4	2	2		1	6	8			8	1	2	1
{45}	3		4				1	3	4		1	7			
{46}	1		5					13	6			7		4	2
{47}	4		2	4			1	4	2	9		8	2	8	2
{48}	1		2				2	3	2			2			
{49}			3	3				2	3	4		5		2	11
{50}	1		3					1		2			5	2	

6

Essay no.												
	SR	SL	WMS	WMM	WRS	WRM	WO	XADJC	XADJPR	XCONJC	XNC	XNPR
{1}	3		1				1					
{2}	4	2	1	1	2	2						
{3}	5		2									
{4}	7		2									
{5}	10	5	2	1	2	2	4					
{6}	7	1	1	1		2	1					
{7}	4				1	1	1					
{8}	3		3	1	2							
{9}	3		1	1			3					
{10}	6					2						
{11}	3	6	2				1					
{12}												
{13}		1			1							
{14}	4				2							
{15}	4	5		1								
{16}	7		1				1					
{17}	5	1					1		2			
{14}	4				2							
{15}	4	5		1								
{16}	7		1				1					
{17}	5	1					1		2			
{18}	2	5	1		1	1						
{19}	6	1			1		1					
{20}												
{21}	5		1		2	1	2					
{23}	11		2		1		2					
{24}	5	8	1	1								
{25}	10				1		1					
{26}	5	4			1		1					
{27}	5	4	2									
{28}	2		2		1							2
{29}	3	1	2		1							
{30}					2							1
{31}	7	3	1	1	1		1					
{33}	5		1									
{34}	3	8	1				5					1
{35}	4		3									
{36}												
{37}			4		3	1						1
{38}	3	1	3			1			1			
{39}	2	1		1	2							
{40}	9						1					

Appendix H

Essay no.												
	SR	SL	WMS	WMM	WRS	WRM	WO	XADJC	XADJPR	XCONJC	XNC	XNPR
{41}	2		2			1						
{42}	4				2							1
{43}	2	7	2						1			
{44}	3	13	7	2	1		1					
{45}	1		1	1	1		1		1			
{46}		3				1	1		1			
{47}	5	4	2	1	2		3					
{48}		1	2		1							
{49}	3	10	2		3		4					
{50}	2						1					

8

Appendix H

Essay no.										
	XNUC	XPRC	XVC	XVPR	Z	GNR	PHU	PHI	GNP	total
{1}			1		3					55
{2)		2	4	2						62
{3}				2	5					87
{4}					3					88
{5}				12	4	1				127
{6}	1				6					87
{7}				3						49
{8}			1	1	1					69
{9}				3						48
{10}				3						56
{11}				3						62
{12}				1	1		1			41
{13)					2					13
{14}	2				3		1			62
{15}										38
{16}	1			3						71
{17}					1					55
{14}	2				3		1			62
{15}										38
{16}	1			3						72
{17}					1					55
{18}		1		2	1					67
{19}	1		1	6	1					82
{20}		1			1					19
{21}				6						67
										0
{23}				2	6					127
{24}				2						46
{25}	1				3					77
{26}				1	4					64
{27}			1	1						26
{28}				1						47
{29}										54
{30}										35
{31}		1	2	1	2					84
										0
{33}	3	1								68
{34}				1						49
{35}	2			2	1					57
{36}				1	2					66
{37}				1	2					53
{38}				3				1		56
{39}			1							72
{40}	1				1					56

9

Appendix H

Essay no.										
	XNUC	XPRC	XVC	XVPR	Z	GNR	PHU	PHI	GNP	total
{41}	4		4							55
{42}										83
{43}										54
{44}				2	1					128
{45}	6			1						75
{46}	2								1	76
{47}	2									115
{48}	2									50
{49}				1						70
{50}										27

10

Appendix I

Polio's (1997) criteria for T units and Clauses attached as they stated in her paper

T-Units

a. A T-unit is defined an independent clause and all its dependent clauses.

b. Count run-on sentences and comma splices as two T-units with an error in the first T-unit.

ex: My school was in Saudi Arabia, it was the best school there.

T / T

1 error error-free

If several comma-splices occur in a row, count only the last as error free.

c. For sentence fragments, if the verb or copula is missing, count the sentence as 1 T-unit with an error. If an NP is standing alone, attach it to the preceding or following T-unit as appropriate and count as an error. If a subordinate clause is standing alone, attach it to the preceding or following S and count it as 1 T-unit with an error.

d. When there is a grammatical subject deletion in a coordinate clause, count the entire sentence as 1 T-unit.

ex: First we went to our school and then went out with our friends.

e. Count both "so" and "but" as coordinating conjunctions. Count "so that" as a subordinating conjunction unless "so" is obviously meant.

f. Do not count tag-questions as separate T-units.

g. Count S-nodes with a deleted complementizer as a subordinate clause as in: I believe that A and (that) B = 1 T-unit.

Appendix I

Polio's (1997) criteria for T units and Clauses attached as they stated in her paper

h. But, direct quotes should be counted as:

John said, "A and B."

1 T-unit 1 T-unit

i. Assess the following type of structures on a case-by-case basis:

If A, then B and C.

As a result, A or B.

j. Count T-units in parentheses as individual T-units.

Clauses

a. A clause equals an overt subject and a finite verb. The following are only one clause each:

He left the house and drove away.

He wanted John to leave the house.

b. Only an imperative does not require a subject to be considered a clause.

c. In a sentence that has a subject with only an auxiliary verb, do not count that subject and verb as a separate clause (or as a separate T-unit. (e.g. John likes to ski and Mary does too; John likes to ski, doesn't he?; John is happy and Mary is too)

Subject Code	TU	EFTU	FDC	FIC	NFC	CT	EFC	EFT/T	EFC/C	EFC/T	Average of three ratios
1A3N	36	20	7	38	11	56	31	0.56	0.55	0.86	0.66
2A3N	24	0	22	25	14	61	8	0.00	0.13	0.33	0.15
3A3N	48	18	21	58	18	97	38	0.38	0.39	0.79	0.52
4A3N	41	5	13	44	16	73	13	0.12	0.18	0.32	0.21
5A3N	58	3	24	59	0	83	5	0.05	0.06	0.09	0.07
6A3N	46	9	33	45	28	106	52	0.20	0.49	1.13	0.61
7A3N	24	1	13	25	12	50	9	0.04	0.18	0.38	0.20
8AN3	22	4	11	24	12	47	7	0.18	0.15	0.32	0.22
9AN3	29	14	20	29	11	60	19	0.48	0.32	0.66	0.48
10A3N	53	19	18	54	7	79	46	0.36	0.58	0.87	0.60
11A3N	38	1	24	43	7	74	8	0.03	0.11	0.21	0.11
12A3N	36	16	26	38	11	75	33	0.44	0.44	0.92	0.60
13A3N	33	21	26	32	14	72	47	0.64	0.65	1.42	0.90
14A3N	39	9	14	43	5	62	14	0.23	0.23	0.36	0.27
15A3N	26	0	11	28	6	45	5	0.00	0.11	0.19	0.10
16A3N	40	5	15	40	5	60	15	0.13	0.25	0.38	0.25
17A3N	27	3	16	27	2	45	13	0.11	0.29	0.48	0.29
18A3N	47	1	20	44	2	66	12	0.02	0.18	0.26	0.15
19A3N	50	9	19	50	6	75	18	0.18	0.24	0.36	0.26
20A3N	27	19	26	28	12	66	61	0.70	0.92	2.26	1.30
						0		######	#####	#DIV/0!	#DIV/0!
58A4N	17	3	7	17	3	27	7	0.18	0.26	0.41	0.28
59A4N	33	9	12	35	7	54	19	0.27	0.35	0.58	0.40
60A4N	37	9	11	38	19	68	37	0.24	0.54	1.00	0.60
61A4N	31	10	13	34	3	50	23	0.32	0.46	0.74	0.51
62A4N	32	14	37	31	5	73	34	0.44	0.47	1.06	0.66
63A4N	43	23	18	41	7	66	36	0.53	0.55	0.84	0.64
64A4N	48	8	18	48	3	69	21	0.17	0.30	0.44	0.30
65A4N	35	6	15	37	22	74	16	0.17	0.22	0.46	0.28
66A4N	33	7	25	33	11	69	21	0.21	0.30	0.64	0.38
67A4N	59	0	24	61	7	92	17	0.00	0.18	0.29	0.16
68A4N	36	2	14	37	2	53	3	0.06	0.06	0.08	0.07
69A4N	32	0	18	32	1	51	4	0.00	0.08	0.13	0.07
70A4N	25	5	13	26	13	52	24	0.20	0.46	0.96	0.54
71A4N	46	10	14	48	7	69	13	0.22	0.19	0.28	0.23
72A4N	28	2	13	29	8	50	9	0.07	0.18	0.32	0.19
73A4N	27	2	48	28	3	79	15	0.07	0.19	0.56	0.27
74A4N	45	1	34	55	10	99	6	0.02	0.06	0.13	0.07
75A4N	40	1	10	42	6	58	15	0.03	0.26	0.38	0.22
76A4N	28	1	32	29	9	70	17	0.04	0.24	0.61	0.30
77A4N	27	6	18	27	2	47	18	0.22	0.38	0.67	0.42
						0		######	#####	#DIV/0!	#DIV/0!
113B3N	19	1	18	20	4	42	4	0.05	0.10	0.21	0.12

114B3N	22	15	7	26	1	34	26	0.68	0.76	1.18	0.88
115B3N	35	9	15	34	0	49	15	0.26	0.31	0.43	0.33
116B3N	29	16	14	29	4	47	19	0.55	0.40	0.66	0.54
117B3N	50	7	11	53	0	64	31	0.14	0.48	0.62	0.41
118B3N	29	6	15	29	9	53	13	0.21	0.25	0.45	0.30
119B3N	16	1	11	19	4	34	10	0.06	0.29	0.63	0.33
120B3N	34	2	11	33	0	44	9	0.06	0.20	0.26	0.18
121B3N	43	8	9	45	4	58	19	0.19	0.33	0.44	0.32
122B3N	45	20	28	45	12	85	43	0.44	0.51	0.96	0.64
123B3N	52	9	17	53	3	73	14	0.17	0.19	0.27	0.21
124B3N	18	10	6	19	3	28	19	0.56	0.68	1.06	0.76
125B3N	53	2	14	57	3	74	19	0.04	0.26	0.36	0.22
126B3N	49	1	5	51	2	58	4	0.02	0.07	0.08	0.06
127B3N	16	5	10	16	4	30	13	0.31	0.43	0.81	0.52
128B3N	30	0	4	29	0	33	2	0.00	0.06	0.07	0.04
129B3N	11	0	6	11	2	19	0	0.00	0.00	0.00	0.00
130B3N	54	4	16	58	2	76	12	0.07	0.16	0.22	0.15
131B3N	21	4	11	21	0	32	7	0.19	0.22	0.33	0.25
132B3N	22	0	10	22	0	32	1	0.00	0.03	0.05	0.03
						0		######	#####	#DIV/0!	#DIV/0!
139B4N	43	10	22	46	2	70	26	0.23	0.37	0.60	0.40
140B4N	32	14	17	35	10	62	29	0.44	0.47	0.91	0.60
141B4N	30	5	14	34	3	51	13	0.17	0.25	0.43	0.28
142B4N	32	0	12	30	5	47	9	0.00	0.19	0.28	0.16
143B4N	30	1	21	33	3	57	9	0.03	0.16	0.30	0.16
144B4N	25	2	13	27	2	42	4	0.08	0.10	0.16	0.11
145B4N	17	3	6	17	2	25	7	0.18	0.28	0.41	0.29
146B4N	20	4	8	21	0	29	14	0.20	0.48	0.70	0.46
147B4N	23	12	11	23	4	38	24	0.52	0.63	1.04	0.73
148B4N	59	16	3	59	0	62	23	0.27	0.37	0.39	0.34
149B4N	57	16	16	58	3	77	28	0.28	0.36	0.49	0.38
150B4N	35	9	23	36	0	59	20	0.26	0.34	0.57	0.39
151B4N	21	4	8	22	1	31	11	0.19	0.35	0.52	0.36
152B4N	26	3	21	27	2	50	9	0.12	0.18	0.35	0.21
153B4N	22	5	18	22	3	43	12	0.23	0.28	0.55	0.35
154B4N	50	5	25	51	4	80	10	0.10	0.13	0.20	0.14
155B4N	22	15	6	21	4	31	22	0.68	0.71	1.00	0.80
156B4N	43	5	15	47	16	78	19	0.12	0.24	0.44	0.27
157B4N	23	4	22	24	3	49	13	0.17	0.27	0.57	0.33
158B4N	45	14	23	46	12	81	32	0.31	0.40	0.71	0.47

Appendix K

Students/ Essay codes	Number of correct sequences (NoCS)	mean length of correct sequences (MLCS)	Column D (NoCS)*(MLCS)	total words in essay	Result of column D/total number of words in essay
1A3N	50	6.56	328	380	0.86
2A3N	63	5.16	325	454	0.72
3A3N	63	6.32	398	524	0.76
4A3N	52	7.54	392	477	0.82
5A3N	84	3.62	304	584	0.52
6A3N	72	6.85	493	577	0.85
7A3N	46	5.15	237	354	0.67
8A3N	44	6.23	274	380	0.72
9A3N	39	9.31	363	453	0.80
10A3N	61	7.62	465	554	0.84
11A3N	74	3.38	250	605	0.41
12A3N	53	8.47	449	504	0.89
13A3N	49	9.84	482	516	0.93
14A3N	54	6.80	367	513	0.72
15A3N	44	4.02	177	312	0.57
16A3N	53	5.57	295	457	0.65
17A3N	51	5.22	266	347	0.77
18A3N	60	4.85	291	500	0.58
19A3N	74	5.72	423	583	0.73
20A3N	33	15.39	508	538	0.94
					#DIV/0!
58A4N	24	9.42	226	275	0.82
59A4N	48	8.29	398	462	0.86
60A4N	46	8.76	403	492	0.82
61A4N	65	5.00	325	419	0.78
62A4N	59	8.66	511	608	0.84
63A4N	57	6.82	389	475	0.82
64A4N	73	5.14	375	535	0.70
65A4N	93	4.80	446	622	0.72
66A4N	57	6.37	363	521	0.70
67A4N	81	5.02	407	708	0.57
68A4N	72	4.13	297	461	0.64
69A4N	60	4.03	242	465	0.52
70A4N	38	9.95	378	408	0.93
71A4N	66	5.08	335	561	0.60
72A4N	50	5.48	274	446	0.61
73A4N	70	6.97	488	633	0.77
74A4N	88	4.01	353	752	0.47
75A4N	49	4.43	217	383	0.57
76A4N	68	6.68	454	531	0.85
77A4N	40	7.88	315	380	0.83

1

Appendix K

Students/ Essay codes	Number of correct sequences (NoCS)	mean length of correct sequences (MLCS)	Column D (NoCS)*(MLCS)	total words in essay	Result of column D/total number of words in essay
113B3N	41	5.83	239	344	0.69
114B3N	22	10.68	235	258	0.91
115B3N	40	7.95	318	401	0.79
116B3N	42	8.43	354	390	0.91
117B3N	65	5.31	345	439	0.79
118B3N	41	5.07	208	335	0.62
119B3N	36	4.67	168	258	0.65
120B3N	37	4.95	183	301	0.61
121B3N	75	5.28	396	578	0.69
122B3N	66	8.12	536	618	0.87
123B3N	70	3.90	273	489	0.56
124B3N	27	7.33	198	214	0.93
125B3N	80	4.46	357	521	0.69
126B3N	67	3.40	228	397	0.57
127B3N	32	5.50	176	221	0.80
128B3N	41	3.07	126	275	0.46
129B3N	21	3.90	82	167	0.49
130B3N	76	3.96	301	530	0.57
131B3N	44	4.32	190	297	0.64
132B3N	40	2.70	108	317	0.34
139B4N	70	5.94	416	528	0.79
140B4N	57	7.42	423	482	0.88
141B4N	51	5.76	294	451	0.65
142B4N	39	3.92	153	272	0.56
143B4N	62	4.39	272	471	0.58
144B4N	51	3.22	164	392	0.42
145B4N	24	6.67	160	201	0.80
146B4N	37	5.95	220	257	0.86
147B4N	27	11.19	302	343	0.88
148B4N	64	5.33	341	483	0.71
149B4N	81	4.58	371	537	0.69
150B4N	53	5.64	299	404	0.74
151B4N	33	6.48	214	279	0.77
152B4N	58	5.19	301	449	0.67
153B4N	39	5.74	224	346	0.65
154B4N	71	4.20	298	581	0.51
155B4N	27	10.81	292	308	0.95
156B4N	81	4.63	375	564	0.66
157B4N	50	4.52	226	351	0.64
158B4N	61	8.13	496	601	0.83

2

Appendix L

Essay no. (9)

It was a week which I met new friends. That week was a worried week because of many new things that I had not seen in the life of school. The first week at college is not like other weeks because there are many things which are unexpected, and sometimes we have in the life of school simple things and simple work. While in the college we face a new style of studying which is more simple than the college style. no one knows how many people are there in their last year of school who are waiting to their first week of college with happiness but the point is how will they enter to the college and what should they do before they enter to it, of course they will first meet new people, new student and teachers. the first week I came to the college was a week full of happiness because I met new friends and new teachers and that week was full of interesting things and happy moments. I felt happy day by day. I was afraid of how to go on with the new style of studying, but now I am going on in college and it is simple for me. on one hand the first week is a week of saddness because of the difficult and the different style of having lectures and many other things but on the other hand it is a week of happiness, first because of getting the college and second because of getting friends and many other things. There are many reasons behind why the first week is important at college and the most important reason is how to go on in a good way to get high marks many people try to get good marks and they do hard in their life and there are people who get accellent degrees, all these point may come from how to be an active one from the first week and how to go on in the following weeks, sometime the first week at college is a normal week like other weeks special for those who are clever and and smart in their life, so they see the firs week as a normal week because the are already had more difficult weeks in their life so those who are on the top, while there are students who see the first everyday and they do not learn new things in their college and those studend see difficulty in their in their life of college, especially those who do not care about knowledge, so the first week at college is a week of difficulties and it is not as simple as normal week and that is it.

Appendix M

My First Week at College

First week at college It different for all students. I thought because student not know more about college, life in college very different for student. So, student saw another life not like life in school because, life in school was very routin, and easy. Studen when came college they not saw that life they saw in school. Maybe student in school had life very nice, and interesting. Students always in school had many friends or best teachers, students in school had very simple study always teacher their family help them not like in college, maybe many students when they came colleg not stayed at home. They farther nom. My first week at college that week was very difficult for me because, when I came college I didn't saw that people I saw in school, and not saw that friends I had in school. When I came college first day when I came my feel is very sad, sorrowful. Because for the first time I far my family and especially my parents my best friends when school always I thought with my best friends. We said, we would go to same college and that tim I didn't think we would seprat each other. I came college. My best friend went another college that's why I didn't saw best friend and my teacher always helpful for me. I didn't know life in college and didn't know the teacher and didn't know students, for the first week that days I didn't have friends because, all students didn't conects each others, that days always thought my family, parents, brothers, and sisters. All days for the first week. I called my father, and mother I said life in college was very difficult and my feel very bad and asked how could stay in college but, when I called my brother. My brother always advised me. He said me, "college was very good life in college students free more than school. All that days when I went upart ment I cried for my family. I missed them. Another day in that first week when I came class I saw some teacher I didn't know them because when I school I know all teacher, and always contected them because some of that teacher lived in our village and I love them so much. My name came in depart English that for the first time when came class. I saw all teachers spoken English were very fast. I didn't understand one words them I thought college was very difficult. I couldn't stay in my feel was very bad, suddenly I thought failed in college I was very worry about all subjects. I could studied them because I couldn't understand them. Day by Day for me become difficult in that first week. I couldn't ask teachers about subjects I always worried always in that time I felt alone nobody near me nobody with me because I didn't know nobody in the college . I saw some students together. My parents always called me and advised me for that feel bad. When I went school I lived with my parents with my brothers, and sisters. When I came home all family helped me family helped me for subjects, food, washing, cleaning, and all worke for me, but I came college I far home nobody had help me nobody work me because I far home, I live Akre not Duhok. I didn't forget all that days for the first week. I couldn't forget them forever.

Appendix N

Do you agree/ disagree that the countryside life is much healthier than the city life?

Healthy is the most happfuly things to make a best life. Good healthy gives us the greate key of clearly futures, so human's depending on his/her environment to be a truth of life. As well as life will be change by the places as human's choice, but most of people say; healthy is a part of environments. I do agree that the countryside life is much healthier than the city life because, countryside has much way to healthy like, environment place and crowded by other ways.

Environments are so influnce of people's healthy, so people by the environments can get best healthy that is to be the way of your wantted to choice. On the other-

ways environment helps you to getting frish air with active body also to be the kye of success in your plans or will be the dreams.

Croweds are the other ways of healthy that is do in the country side, because as we see crowed in citys are most than the country side. As well as, city croweded has influenced healthy and your plans because, your plans related your health. most people thinks leave the city by the croweded ways also government's tried for keeping it like, factories, cars, also make some other buildings in outsides but the main reseans of crowed is people.

Tonicity is the most greatfuly of things to make frish life. the best way of tonicity to create the big kye of accually fiuture. Also people depending on his/her environments to be truth of life. Tonicity changed by the place that is the frish air like which place has more tries. Also accepts countryside life relate with your healthier as they have better frish air and water.

Appendix O

Your first week at College

College is a place where people get educated, meet new friends and share information together. My first week at College, I felt something new in my personal life because I could meet different people from various areas and I felt really happy about it, at the same time I was somehow scared how to deal with them due to their personalities and behaviours. I could do great things in my first week of College like preparing presentations and Collaborating the other friends was something exciting. I tried to choose some people at College and make them close-friends according to my own personality, everyone couldn't not be a friend of my occupation first week at College, I was missing my family as well, because it was my first time to get far away from my mother, father and brothers, I realized how tough their job was, so I tried to be punctual and serious to my studying because my father over-worked to get money for the sake of me and I highly appreciated his fatigue. When I first came to College my hobby was that to be a brilliant translator and make my dream come true. My friends were very helpful with me at the first days at College and I didn't even feel that I am forigen among then because whenever I had difficulties with some topics, they would be very kind and help me without taking any excuses. In my first days at aPartment, we used to have much fun and work together in every single job, we tried to cock some food and learn from each other, actually it was really interesting because sometimes we burnt the food and some other times it was very delicious. My Partners of aPartment were from different places and each one had his own dialect, So at first their dialect was not really clear for me but after being together for a long time I could deal with their own dialects.

In conclusion, College to me was an impressive place that taught me how to deal with sensitive, serious and funny people and I learnt many things from my teachers, friends and even my personal mistakes, eventually I realized how to live my life and make us of it. Although meeting various friends is not a matter of exploitation or having fun all the time, but it is a process of cooperation and helping each other in all situations

Appendix P

Student codes	CCT	NCCT only punctution error at sentence boundary	NCCT	No. of errors	average of error per NCCTs	UT	CCS	NCCS
1A3N	20	0	16	21	1.31	0	0	0
2A3N	0	0	19	42	2.21	5	14	3
3A3N	17	4	30	60	2.00	0	0	0
4A3N	4	15	30	49	1.63	5	2	6
5A3N	4	20	34	70	2.06	19	25	21
6A3N	9	12	29	49	1.69	3	8	0
7A3N	1	5	19	27	1.42	2	2	3
8AN3	4	5	12	23	1.92	4	15	3
9AN3	14	8	12	17	1.42	2	3	0
10A3N	16	15	29	41	1.41	0	0	0
11A3N	1	17	13	25	1.92	18	36	28
12A3N	16	0	20	39	1.95	0	0	0
13A3N	21	5	9	16	1.78	0	0	0
14A3N	9	13	28	36	1.29	1	4	0
15A3N	0	18	10	17	1.70	13	17	5
16A3N	3	16	27	50	1.85	5	5	4
17A3N	2	12	17	40	2.35	2	4	0
18A3N	1	29	27	46	1.70	6	14	5
19A3N	8	10	40	61	1.53	0	0	0
20A3N	19	1	5	5	1.00	0	0	0
58A4N	3	4	12	25	2.08	0	0	0
59A4N	9	7	21	31	1.48	1	1	3
60A4N	9	10	21	31	1.48	3	5	3
61A4N	9	9	17	34	2.00	1	4	2
62A4N	11	0	20	29	1.45	0	0	0
63A4N	23	7	17	22	1.29	0	0	0
64A4N	7	20	36	57	1.58	0	0	0
65A4N	6	9	17	31	1.82	9	24	12
66A4N	7	11	18	30	1.67	6	13	8
67A4N	0	45	39	69	1.77	11	28	7
68A4N	2	26	20	35	1.75	12	11	14
69A4N	0	14	20	47	2.35	9	12	17
70A4N	4	9	15	21	1.40	1	2	0
71A4N	9	10	36	80	2.22	1	3	0
72A4N	2	13	18	33	1.83	5	6	8
73A4N	2	3	22	49	2.23	1	3	1
74A4N	0	19	29	58	2.00	17	24	11
75A4N	1	17	33	54	1.64	2	4	1
76A4N	1	8	20	33	1.65	1	1	2
77A4N	6	9	17	26	1.53	1	3	1

1

Appendix P

Student codes	CCT	NCCT only punctution error at sentence boundary	NCCT	No. of errors	average of error per NCCTs	UT	CCS	NCCS
113B3N	1	1	17	39	2.29	1	3	2
114B3N	14	2	7	10	1.43	1	2	0
115B3N	9	1	26	40	1.54	1	2	2
116B3N	14	2	14	25	1.79	0	0	0
117B3N	7	24	31	51	1.65	1	2	0
118B3N	6	9	17	35	2.06	6	5	6
119B3N	1	5	11	25	2.27	2		6
120B3N	2	13	25	50	2.00	1	0	2
121B3N	5	12	30	63	2.10	5	4	9
122B3N	19	1	26	28	1.08	0	0	0
123B3N	9	13	31	57	1.84	7	8	3
124B3N	10	0	8	10	1.25	0	0	0
125B3N	2	28	36	66	1.83	7	8	9
126B3N	1	24	40	67	1.68	4	5	1
127B3N	5	0	11	19	1.73	0	0	0
128B3N	0	6	27	63	2.33	2	0	0
129B3N	0	3	7	21	3.00	3	4	2
130B3N	4	26	42	86	2.05	4	5	3
131B3N	3	1	17	30	1.76	1	2	4
132B3N	0	1	5	8	1.60	16	30	5
139B4N	10	13	29	40	1.38	0	0	0
140B4N	14	2	16	24	1.50	0	0	0
141B4N	5	11	16	34	2.13	4	12	5
142B4N	0	12	27	53	1.96	2	0	1
143B4N	1	12	19	48	2.53	5	11	2
144B4N	2	6	10	20	2.00	13	16	11
145B4N	3	4	11	21	1.91	0	0	0
146B4N	4	12	10	24	2.40	0	0	0
147B4N	11	1	10	15	1.50	0	0	0
148B4N	16	21	35	50	1.43	1	3	0
149B4N	16	12	36	57	1.58	0	0	0
150B4N	8	7	22	41	1.86	2	3	2
151B4N	4	7	12	23	1.92	1	3	0
152B4N	3	12	15	34	2.27	4	7	7
153B4N	4	1	14	27	1.93	3	5	5
154B4N	4	17	37	78	2.11	6	8	8
155B4N	15	1	6	10	1.67	0	0	0
156B4N	5	12	33	66	2.00	3	3	5
157B4N	4	7	15	32	2.13	2	7	3
158B4N	12	11	27	35	1.30	1	3	2

Appendix P

Student codes	CCS/UT	NCCS/UT	average length of CCS	Average length of NCCS	all T units	Fragments but still well formed
1A3N	0.00	0.00			36	
2A3N	2.80	0.60	3.92	5.33	24	
3A3N	0.00	0.00			48	4
4A3N	0.40	1.20	8	7.83	41	
5A3N	1.32	1.11	3.2	3.55	58	2
6A3N	2.67	0.00	7.37		46	3
7A3N	1.00	1.50	4	6	24	1
8AN3	3.75	0.75	3.8	4	22	
9AN3	1.50	0.00	3		29	
10A3N	0.00	0.00			53	
11A3N	2.00	1.56	3.5	4	38	
12A3N	0.00	0.00			36	
13A3N	0.00	0.00			33	
14A3N	4.00	0.00	4.25		39	
15A3N	1.31	0.38	4.76	5.6	26	1
16A3N	1.00	0.80	4.8	3.5	40	
17A3N	2.00	0.00	2.25		27	
18A3N	2.33	0.83	4.28	6.6	47	
19A3N	0.00	0.00			50	
20A3N	0.00	0.00			27	3
58A4N					17	
59A4N					33	
60A4N					37	
61A4N					31	
62A4N					32	
63A4N					43	
64A4N					48	
65A4N					35	
66A4N					33	
67A4N					59	
68A4N					36	
69A4N					32	
70A4N					25	
71A4N					46	2
72A4N					28	
73A4N					27	
74A4N					45	
75A4N					40	
76A4N					28	
77A4N					27	

3

Appendix P

Student codes	CCS/UT	NCCS/UT	average length of CCS	Average length of NCCS	all T units	Fragments but still well formed
113B3N					19	
114B3N					22	
115B3N					35	
116B3N					29	
117B3N					50	
118B3N					29	
119B3N					16	1
120B3N					34	
121B3N					43	
122B3N					45	
123B3N					52	
124B3N					18	
125B3N					53	
126B3N					49	
127B3N					16	
128B3N					30	
129B3N					11	
130B3N					54	
131B3N					21	
132B3N					22	
139B4N					43	
140B4N					32	
141B4N					30	
142B4N					32	
143B4N					30	
144B4N					25	
145B4N					17	
146B4N					20	
147B4N					23	
148B4N					59	
149B4N					57	1
150B4N					35	
151B4N					21	
152B4N					26	
153B4N					22	
154B4N					50	
155B4N					22	
156B4N					43	
157B4N					23	
158B4N					45	

4

Student codes	fragments not well formed	CCT/T	NCCT-P/T	NCCT/T	UT/T	
1A3N		0.56	0.00	0.44	0.00	
2A3N		0.00	0.00	0.79	0.21	
3A3N		0.35	0.08	0.63	0.00	
4A3N	1	0.10	0.37	0.73	0.12	
5A3N	4	0.07	0.34	0.59	0.33	
6A3N		0.20	0.26	0.63	0.07	
7A3N		0.04	0.21	0.79	0.08	
8AN3		0.18	0.23	0.55	0.18	
9AN3		0.48	0.28	0.41	0.07	
10A3N	1	0.30	0.28	0.55	0.00	
11A3N		0.03	0.45	0.34	0.47	
12A3N	1	0.44	0.00	0.56	0.00	
13A3N		0.64	0.15	0.27	0.00	
14A3N		0.23	0.33	0.72	0.03	
15A3N		0.00	0.69	0.38	0.50	
16A3N	2	0.08	0.40	0.68	0.13	
17A3N		0.07	0.44	0.63	0.07	
18A3N		0.02	0.62	0.57	0.13	
19A3N	1	0.16	0.20	0.80	0.00	
20A3N		0.70	0.04	0.19	0.00	
58A4N		0.18	0.24	0.71	0.00	
59A4N		0.27	0.21	0.64	0.03	
60A4N		0.24	0.27	0.57	0.08	
61A4N		0.29	0.29	0.55	0.03	
62A4N		0.34	0.00	0.63	0.00	
63A4N		0.53	0.16	0.40	0.00	
64A4N		0.15	0.42	0.75	0.00	
65A4N		0.17	0.26	0.49	0.26	
66A4N		0.21	0.33	0.55	0.18	
67A4N		0.00	0.76	0.66	0.19	
68A4N		0.06	0.72	0.56	0.33	
69A4N		0.00	0.44	0.63	0.28	
70A4N		0.16	0.36	0.60	0.04	
71A4N		0.20	0.22	0.78	0.02	
72A4N		0.07	0.46	0.64	0.18	
73A4N		0.07	0.11	0.81	0.04	
74A4N		0.00	0.42	0.64	0.38	
75A4N		0.03	0.43	0.83	0.05	
76A4N		0.04	0.29	0.71	0.04	
77A4N		0.22	0.33	0.63	0.04	

5

Appendix P

Student codes	fragments not well formed	CCT/T	NCCT-P/T	NCCT/T	UT/T	
113B3N		0.05	0.05	0.89	0.05	
114B3N		0.64	0.09	0.32	0.05	
115B3N		0.26	0.03	0.74	0.03	
116B3N		0.48	0.07	0.48	0.00	
117B3N		0.14	0.48	0.62	0.02	
118B3N		0.21	0.31	0.59	0.21	
119B3N	1	0.06	0.31	0.69	0.13	
120B3N		0.06	0.38	0.74	0.03	
121B3N		0.12	0.28	0.70	0.12	
122B3N		0.42	0.02	0.58	0.00	
123B3N	5	0.17	0.25	0.60	0.13	
124B3N		0.56	0.00	0.44	0.00	
125B3N		0.04	0.53	0.68	0.13	
126B3N		0.02	0.49	0.82	0.08	
127B3N		0.31	0.00	0.69	0.00	
128B3N	1	0.00	0.20	0.90	0.07	
129B3N		0.00	0.27	0.64	0.27	
130B3N	1	0.07	0.48	0.78	0.07	
131B3N		0.14	0.05	0.81	0.05	
132B3N		0.00	0.05	0.23	0.73	
139B4N		0.23	0.30	0.67	0.00	
140B4N		0.44	0.06	0.50	0.00	
141B4N		0.17	0.37	0.53	0.13	
142B4N		0.00	0.38	0.84	0.06	
143B4N		0.03	0.40	0.63	0.17	
144B4N	1	0.08	0.24	0.40	0.52	
145B4N		0.18	0.24	0.65	0.00	
146B4N		0.20	0.60	0.50	0.00	
147B4N		0.48	0.04	0.43	0.00	
148B4N		0.27	0.36	0.59	0.02	
149B4N	2	0.28	0.21	0.63	0.00	
150B4N		0.23	0.20	0.63	0.06	
151B4N		0.19	0.33	0.57	0.05	
152B4N		0.12	0.46	0.58	0.15	
153B4N		0.18	0.05	0.64	0.14	
154B4N		0.08	0.34	0.74	0.12	
155B4N		0.68	0.05	0.27	0.00	
156B4N	2	0.12	0.28	0.77	0.07	
157B4N		0.17	0.30	0.65	0.09	
158B4N		0.27	0.24	0.60	0.02	

Appendix Q

e.g. 8.51 essay (37) attached as

Country Side Vs. City life

Every people in this world live in different places and each one of them prefer some places to live in. The old people who lived in the past, they live in countryside because there was no city life on that time. Nowadays the life become very modern. Every individual in this world prefer countryside or City life. In my point of view, City life is better than the countryside because it's the time of technology and in every step in your life, we will need technological tools and things like internet, mobile, and so on. Also there are many places in city which we can go and enjoy with our family and friends such as parks, cofees, libraries and many other places that is not exist in countryside. Of course there are many drawbacks about cities like existence of many factories and places that create excessive waste and pollution. This results are the main reason to pollute the enviroment which is very bad and unhealthy for us, but beside this disadvantages of City life, we can not live in countryside because it's became as a habit for us and we can not live without this things that we have in City life. Also life in City is more easier than the countryside because there are many things in City that help us to live easily such as many tools, machines but in countryside, we have to do every thing by ourselves, and it is very difficult for us because our generation in nowadays are not as strong as the old ones. Finally we can say that the city life is a better and easier than the countryside to live in it and also to have a comfortable life.

References

Abdi, M., Eslami, H. & Zahedi, Y. 2012. The impact of pre-task planning on the fluency and accuracy of Iranian EFL learners' oral performance. *Procedia-Social and Behavioral Sciences*. 69, pp. 2281–2288.

Abdollahifam, S. 2014. Investigating the effects of interactional feedback on EFL students' writings. *Procedia-Social and Behavioral Sciences*. 98, pp. 16–21.

Ädel, A. & Erman, B. 2012. Recurrent word combinations in academic writing by native and non-native speakers of English: A lexical bundles approach. *English for Specific Purposes*. 31(2), pp. 81–92.

Ahmadian, M.J. 2012. The effects of guided careful online planning on complexity, accuracy and fluency in intermediate EFL learners' oral production: The case of English articles. *Language Teaching Research*. 16(1), pp. 129-149.

Ahmadian, M.J., Tavakoli, M. & Dastjerdi, H.V., 2015. The combined effects of online planning and task structure on complexity, accuracy and fluency of L2 speech. *The Language Learning Journal*. 43(1), pp. 41–56.

Ahmadian, M.J. & Tavakoli, M. 2010. The effects of simultaneous use of careful online planning and task repetition on accuracy, complexity, and fluency in EFL learners' oral production. *Language Teaching Research*. 15(1), pp. 35–59.

Ai, H. and Lu, X. 2013. 'A Corpus-Based Comparison of Syntactic Complexity in NNS and NS University Students' Writing'. In A. Díaz-Negrillo, N. Ballier, & P. Thompson (eds). *Automatic Treatment and Analysis of Learner Corpus Data*. Amsterdam: John Benjamins, pp. 249–264.

Akbarzadeh, R., Saeidi, M., & Chehreh, M. 2014. The effect of oral interactive feedback on the accuracy and complexity of EFL learners' writing performance: Uptake and retention. *Iranian Journal of Language Teaching Research*. 2(2), pp. 105–126.

Altenberg, B. & Tapper, M. 1998. 'The Use of Adverbial Connectors in Advanced Swedish Learners' Written English'. In S. Granger (ed.) *Learner English on Computer*. New York: Addison Wesley Longman, pp. 80–93.

Arnaud, P.J.L. 1992. 'Objective Lexical and Grammatical Characteristics of L2 Written Compositions and the Validity of Separate-Component Tests'. In P.J.L. Arnaud & H. Bejoint (eds). *Vocabulary and Applied Linguistics*. London: Macmillan, pp. 133–145.

Arslanyilmaz, A., & Pedersen, S. 2010. Improving language production using subtitled similar task videos. *Language Teaching Research*. 14(4), pp. 377–395.

Arthur, B. 1979. 'Short-Term Changes in EFL Compositions Skills'. In C. Yorio, K. Perkins, & J. Schachter (eds). *On TESOL 79: The learner in Focus*. Washington D. C.: TESOL, pp. 330–342.

Ashwell, T. 2000. Patterns of teacher response to student writing in a multiple-draft composition classroom: Is content feedback followed by form feedback the best method? *Journal of Second Language Writing*. 9(3), pp. 227–257.

Azodi, N., Karimi, F. & Vaezi, R. 2014. Measuring the lexical richness of productive vocabulary in Iranian EFL university students' writing performance. *Theory and Practice in Language Studies*. 4(9), pp. 1837–1849.

Bagheri, M. & Hamrang, A. 2013. The impact of planning on accuracy and complexity in oral production of male and female English as a foreign language (EFL) learners. *International Journal of Applied Linguistics and English Literature*. 2(2), pp. 25–32.

Bamanger, E. M. & Gashan, A. K. 2015. The effect of planning time on the fluency, accuracy, and complexity of EFL learners' oral production. *Journal of Educational Sciences*. 27(1), pp. 1–15.

Bardovi-Harlig, K. 1992. A second look at T-unit analysis: Reconsidering the sentence. *TESOL quarterly*. 26(2), pp. 390–395.

Bardovi-Harlig, K. & Bofman, T. 1989. Attainment of syntactic and morphological accuracy by advanced language learners. *Studies in Second Language Acquisition*. 11(01), pp. 17–34.

Barkaoui, K. 2010. Explaining ESL essay holistic scores: A multilevel modeling approach. *Language Testing*. 27(4), pp. 515–535.

Biber, D., Gray, B. & Poonpon, K. 2011. Should we use characteristics of conversation to measure grammatical complexity in L2 writing development? *Tesol Quarterly*. 45(1), pp. 5–35.

Bitchener, J., Young, S. & Cameron, D. 2005. The effect of different types of corrective feedback on ESL student writing. *Journal of Second Language Writing*. 14(3), pp. 191–205.

Bitchener, J. 2008. Evidence in support of written corrective feedback. *Journal of Second Language Writing*. 17(2), pp. 102–118.

Bitchener, J., & Knoch, U. 2008. The value of written corrective feedback for migrant and international students. *Language Teaching Research*. 12(3), pp. 409–431.

Bitchener, J., & Knoch, U. 2010. Raising the linguistic accuracy level of advanced L2 writers with written corrective feedback. *Journal of Second Language Writing*, 19(4). pp. 207–217.

Bolton, K., Nelson, G. & Hung, J. 2002. A corpus- based study of connectors in student writing: Research from the International Corpus of English in Hong Kong (ICE-HK). *International Journal of Corpus Linguistics*. 7(2), pp. 165–182.

Brock, C.A. 1986. The effects of referential questions on ESL classroom discourse. *TESOL quarterly*. 20(1), pp. 47–59.

Brown, R. 1973. *A First Language: The Early Stages*. Cambridge: Harvard University Press.

Brown, H. D. 2007. *Principles of Language Learning and Teaching.* 4th edn. New York: Longman.

Brumfit, C.J. 1984. *Communicative Methodology in Language Teaching*. Cambridge: Cambridge University Press.

Bulté, B. & Housen, A. 2014. Conceptualizing and measuring short-term changes in L2 writing complexity. *Journal of Second Language Writing*. 26, pp. 42–65.

Bulté, B. & Housen, A 2012. 'Defining and Operationalizing L2 Complexity'. In *Dimensions of L2 Performance and Proficiency Complexity, Accuracy and Fluency in SLA*. Philadelphia: John Benjamins, pp. 21-46.

Burt, M.K., & Kiparsky, C. 1974. 'Global and Local Mistakes'. In J.H. Schumann and N. Stenson (eds). *New Frontiers In Second Language Learning*. Rowley: Newbury House Publishers, pp. 71–80.

Butler, Y.G. 2002. Second language learners' theories on the use of English articles. *Studies in Second Language Acquisition*. 24(03), pp. 451–480.

Byrnes, H., & Manchn, R. M. (eds). 2014. *Task-Based Language Learning: Insights From and for L2 Writing*. Amsterdam: John Benjamins.

Carroll, J.B. 1967. 'On Sampling from a Lognormal Model of Word – Frequency Distribution'. In H. Kucera & W. N. Francis (eds). *Computational Analysis of Present-Day American English*. Providence: Brown University, pp. 406-424.

Casanave, C.P., 1994. Language development in students' journals. *Journal of Second Language Writing*, 3(3), pp.179–201.

Chan, A.Y.W. 2010. Toward a taxonomy of written errors: Investigation into the written errors of Hong Kong Cantonese ESL learners. *TESOL Quarterly*. 44(2), pp. 295–319.

Chandler, J. 2003. The efficacy of various kinds of error feedback for improvement in the accuracy and fluency of L2 student writing. *Journal of Second Language Writing*. 12(3), pp. 267–296.

Chaudron, C., & Parker, K. 1990. Discourse markedness and structural markedness. The acquisition of English noun phrases. *Studies in Second Language Acquisition*.12, pp. 43-64.

Chen, M. 2013. Overuse or underuse: A corpus study of English phrasal verb use by Chinese, British and American university students. *International Journal of Corpus Linguistics*. 18(3), pp. 418–442.

Cheng, W. 2012. *Exploring Corpus Linguistics: Language in Action.* Abingdon:Routledge.

Chu, R. 2011. Effects of teacher's corrective feedback on accuracy in the oral English of English-majors college students. *Theory and Practice in Language Studies*. 1(5), pp. 454–459.

Corder, S.P. 1967. The significance of learner's errors. *IRAL-International Review of Applied Linguistics in Language Teaching*. 5(1–4), pp. 161–170.

Corder, S.P. 1975. Error analysis, interlanguage and second language acquisition. *Language Teaching*. 8(04), pp. 201–218.

Corder, S.P. 1981. *Error analysis and interlanguage* (Vol. 112). Oxford: Oxford University Press.

Crowhurst, M. 1987. Cohesion in argument and narration at three grade levels. *Research in the teaching of English.* 21(2), pp. 185–201.

Cumming, A., & Mellow, D. 1996. 'An Investigation into The Validity of Written Indicators of Second Language Proficiency'. In A. Cumming & R. Berwick (eds). *Validation in Language Testing*. Clevedon: Multilingual Matters, pp. 72–93.

Czwenar, I. 2014. 'Analyzing Spoken Language for Complexity, Accuracy and Fluency: Some Methodological Considerations'. In W. Szubko-Sitarek, L.Salski, & P. Stalmaszczyk, (eds). *Language Learning, Discourse and Communication: Studies in Honour of Jan Majer*. London: Springer Science & Business Media, pp. 81–92.

Dagneaux, E.S. Denness, S. Granger and F. Meunier 1996. Error Tagging Manual, Version 1.1. Center for English Corpus Linguistics. Louvain-la-Neuve: Université Catholique de Louvain.

Dagneaux, E., Denness, S., & Granger, S. 1998. Computer-aided error analysis. *System*. 26(2): pp. 163–174.

Daller, H., Roeland V.H. & Jeanine T.-D. 2003. Lexical richness in the spontaneous speech of bilinguals. *Applied Linguistics*. 24(2), pp. 197–222.

De Bot, K., Lowie, W., & Verspoor, M. 2005. *Second Language Acquisition: An Advanced Resource Book*. New York: Routledge.

Díaz-Negrillo, A. & Domínguez, J.F., 2006. Error tagging systems for learner corpora. *Revista española de lingüística aplicada*, (19), pp.83-102.

Djigunović, J.M. & Krevelj, S.L. 2011. Lexical diversity in early learning of English as L2. *Studia Romanica et Anglica Zagrabiensia*. 56, pp. 253–270.

Dulay H., Burt M. & Krashen St. 1982. *Language Two*. Oxford: Oxford UniversityPress.

Dutra, D.P. & Sardinha, T.B. 2013. 'Referential Expressions in English Learner Argumentative Writing'. In S. Granger, G. Gilquin & F. Meunier (eds). *Twenty Years of Learner Corpus Research: Looking Back, Moving Ahead*. Louvain-La-Neuve: Universitaires de Louvain, pp. 117–127.

Ellis, N.C. 2008. 'Usage-Based and Form-Focused Language Acquisition: The Associative Learning of Constructions, Learned Attention, and The Limited L2 Endstate'. In N. Ellis & P. Robinson (eds). *Handbook of Cognitive Linguistics and Second Language Acquisition*. London: Routledge, pp. 372–405.

Ellis, R. 1994. *The Study of Second Language Acquisition*. 1st edn. Oxford: Oxford University Press.

Ellis, R. 2003. *Task-based Language Learning and Teaching*. Oxford: Oxford University Press.

Ellis, R. 2008. *The Study of Second Language Acquisition*. 2nd edn. Oxford: Oxford University Press.

Ellis, R. & Barkhuizen, G. 2005. *Analysing Learner Language*. Oxford: Oxford University Press.

Ellis, R. Sheen, Y., Murakami, M. & Takashima, H. 2008. The effects of focused and unfocused written corrective feedback in an English as a foreign language context. *System.* 36(3), pp. 353–371.

Ellis, R. & Yuan, F. 2004. The effects of planning on fluency, complexity, and accuracy in second language narrative writing. *Studies in Second Language Acquisition.* 26(01), pp. 59–84.

Engber, C.A. 1995. The relationship of lexical proficiency to the quality of ESL compositions. *Journal of second language writing.* 4(2), pp. 139–155.

Evans, M., & Fisher, L. 2005. Measuring gains in pupils' foreign language competence as a result of participation in a school exchange visit: The case of Y9 pupils at three comprehensive schools in the UK. *Language Teaching Research.* 9(2), pp. 173–192.

Evans, N.W., Hartshorn, K. J., McCollum, R., & Wolfersberger, M. 2010. Contextualizing corrective feedback in second language writing pedagogy. *Language Teaching Research.* 14(4), pp. 445–463.

Evans, N.W., Hartshorn, K.J. and Strong-Krause, D. 2011. The efficacy of dynamic written corrective feedback for university-matriculated ESL learners. *System.* 39(2), pp. 229–239.

Evans, N.W., Hartshorn, K.J., Cox, T.L. & De Jel, T.M. 2014. Measuring written linguistic accuracy with weighted clause ratios: A question of validity. *Journal of Second Language Writing*. 24, pp. 33–50.

Evola, J., Mamer, E. & Lentz, B. 1980. 'Discrete Points Versus Global Scoring for Cohesive Devices. In J. W. Oller & K. Perkins (eds). *Research in Language Testing*. Rowley, MA: Newbury House, pp. 177–181.

Farrokhi, F. & Sattarpour, S. 2011. The Effects of focused and unfocused written corrective feedback on grammatical accuracy of Iranian EFL learners. *Theory and Practice in Language Studies*. 1(12), pp. 1797–1803.

Fazio, L.L. 2001. The effect of corrections and commentaries on the journal writing accuracy of minority-and majority-language students. *Journal of Second Language Writing*. 10(4), pp. 235–249.

Fischer, R.A. 1984. Testing written communicative competence in French. *Modern Language*

Journal. 68 (1), pp. 13–20.

Flahive, D. E. & Snow, B. G. 1980. 'Measures of Syntactic Complexity in Evaluating ESL compositions'. In J. W. Oller & K. Perkins (eds). Research in Language Testing. Rowley, MA: Newbury House, pp. 171–176.

Foster, P. & Skehan, P. 1996. The influence of planning and task type on second language performance. *Studies in Second Language Acquisition*. 18(03), pp. 299–323.

Foster, P. and Skehan, P. 1999. The influence of source of planning and focus of planning on task-based performance. *Language Teaching Research*. 3(3), pp. 215–247.

Foster, P., Tonkyn, A. & Wigglesworth, G. 2000. Measuring spoken language: A unit for all reasons. *Applied Linguistics*. 21(3), pp. 354–375.

Frantzen, D. 1995. The effects of grammar supplementation on written accuracy in an intermediate Spanish content course. *The Modern Language Journal*. 79(3), pp. 329–344.

Freed, B.F. 1978. *Foreigner talk: A study of speech adjustments made by native speakers of English in conversation with non-native speakers*. Unpublished doctoral dissertation, University of Pennsylvania, Philadelphia.

Gaies, S.J. 1980. T-unit analysis in second language research: Applications, problems and limitations. *TESOL quarterly*. 14(1), pp. 53–60.

Gass, S., & Selinker, L. 2001. *Second Language Acquisition: An Introductory Course.* 2nd edn. New York: Taylor & Francis.

Gass, S., & Selinker, L. 2008. *Second Language Acquisition: An Introductory Course*. 3rd edn. New York: Taylor & Francis.

Ghavamnia, M., Tavakoli, M. & Esteki, M. 2013. The effect of pre-task and online planning conditions on complexity, accuracy, and fluency on EFL learners written production. *Porta Linguarum: Revista Internacional de Didáctica de Las Lenguas Extranjeras* (20), pp. 31–43.

Gholizade, R. 2013. The investigation of differential effects of recast and metalinguistic feedback on accuracy, fluency, and complexity of speaking performance of male and female EFL learners. *Journal of Novel Applied Sciences.* 2 (9), pp. 417–428.

Ghonsooly, B., & Shalchy, S. 2013. Cultural intelligence and writing ability: delving into fluency, accuracy and complexity. *Novitas-ROYAL-Research on Youth and Language*. 7(2), pp. 147–159.

Grabe, W. 1985. 'Written Discourse Analysis'. In R.B. Kaplan, A. d'Anglejan, J.R. Cowan, B. Kachru, G.R. Tucker, & H. Widdowson (eds). *Annual Review of Applied Linguistics* Vol. 5. New York: Cambridge University Press, pp. 101–123.

Granger, S. 1996. From ca to cia and back: An integrated approach to computerized bilingual and learner corpora. In K. Aijmer, B. Altenberg, & M. Johansson (eds.), Languages in contrast: Paper from a symposium on text-based cross-linguistic studies. Lund studies in English Vol. 88. Lund: Lund University Press, pp. 37–51.

Granger, S. (ed.) 1998a. *Learner English on Computer*. London: Longman.

Granger, S. 1998b. 'Prefabricated Patterns in Advanced EFL Writing: Collocations and Formulae'. In A.P. Cowie (ed.) *Phraseology: Theory, Analysis, and Applications*. Oxford: Oxford University Press, pp. 145–160.

Granger, S. 2002. 'A Bird's-Eye View of Learner Corpus Research'. In G. Sylviane, J. Hung & S. Petch Tyson (eds). *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*. Amsterdam: John Benjamins, pp. 3–33.

Granger, S. 2003. Error-tagged learner corpora and CALL: a promising synergy. *CALICO Journal,* 20 (3). Special issue on error analysis and error correction in computer-assisted language learning: pp. 465-480.

Granger, S. 2004. 'Computer Learner Corpus Research: Current Status And Future Prospects'. In U. Connor (ed.) *Applied Corpus Linguistics: A Multidimensional Perspective*. Amsterdam: Rodopi, pp. 123–145.

Granger, S. 2008. 'Learner Corpora'. In A. Lüdeling and M. Kytö (eds). *Corpus Linguistics—An International Handbook.* Vol. 1. Berlin: Walter de Gruyter, pp. 259–274.

Granger, S. 2009. 'The Contribution of Learner Corpora To Second Language Acquisition And Foreign Language Teaching: A Critical Evaluation'. In K. Ajmer (ed.) *Corpora and Language Teaching.* Amsterdam: Benjamins, pp. 13–32.

Granger, S. & Tyson, S. 1996. Connector usage in the English essay writing of native and non-native EFL speakers of English. *Word Englishes*. 15(1), pp. 17– 27.

Granger, S., & Meunier, F. (eds) 2008. 'Phraseology in Language Learning and Teaching: Where to from Here?' In *Phraseology in Foreign Language Learning and Teaching*. Amsterdam: John Benjamins, pp. 247–252.

Granger, S., Dagneaux, E., Meunier, F., & Paquot, M. (2009). International corpus of learner English (Version 2.0). Louvain-la-Neuve: Presses Universitaires de Louvain.

Halliday, M.A.K. 1985. *Spoken and Written Language.* Geelong Vict: Deakin University.

Halliday, M. and Martin, J. 1993. *Writing Science: Literacy and Discursive Power.* London: Falmer.

Halliday, M. and Matthiessen C. 1999. *Construing Experience through Meaning: A Language-Based Approach to Cognition*. London: Continuum.

Han, Zh. 2004. *Fossilization in Adult Second Language Acquisition*. Clevedon: Multilingual Matters.

Harley, B. & King, M.L. 1989. Verb lexis in the written compositions of young L2 learners. *Studies in Second Language Acquisition*. 11(04), pp. 415–439.

Hasselgård, H. & Johansson, S. 2011. 'Learner Corpora and Contrastive Interlanguage Analysis'. In F. Meunier, S. De Cock, G. Gilquin and M. Paquot (eds). *A Taste for*

Corpora: In Honour of Sylviane Granger. Amsterdam & Philadelphia: John Benjamins, pp. 33–62.

Homburg, T.J. 1984. Holistic evaluation of ESL compositions: Can it be validated objectively? *TESOL quarterly*. 18(1), pp. 87–107.

Ho-Peng, L. 1983. Using T-unit measures to assess writing proficiency of university ESL students. *RELC Journal*. 14(2), pp. 35–43.

Housen, A. 2002. 'A Corpus-Based Study of the L2-Acquisition of the English Verb System'. In S. Granger, J. Hung, & S. Petch-Tyson (eds). *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*. Amsterdam & Philadelphia: John Benjamins, pp. 77–116.

Housen, A. & Kuiken, F. 2009. Complexity, accuracy, and fluency in second language acquisition. *Applied linguistics*. 30(4), pp. 461–473.

Housen, A., Kuiken, F. & Vedder, I. (eds) 2012. 'Complexity, Accuracy and Fluency: Definitions, Measurement and Research'. In *Dimensions of L2 Performance and Proficiency Complexity, Accuracy and Fluency in SLA*. Philadelphia: John Benjamins, pp. 1–20.

Howarth, P.A. 1998. 'The Phraseology of Learners' Academic Writing'. In A.P. Cowie (ed.) *Phraseology: Theory, Analysis, and Applications*. Oxford: Oxford University Press, Pp. 160–186.

Huddleston, R. & Pullum, G.K. (2005) *A Student's Introduction to English Grammar*. Cambridge: Cambridge University Press.

Hunston, S. 2002. *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.

Hunston, S. 2008. 'Collection Strategies and Design Decisions'. In A. Lüdeling and M. Kytö (eds). *Corpus Linguistics—An International Handbook.* Vol. 1. Berlin: Walter de Gruyter, pp. 154–168.

Hunt, K.W. 1965. *Grammatical structures written at three grade levels*. NCTE Research Report No. 3. Champaign, IL: National Council of Teachers of English.

Hunt, K.W. 1966. Recent measures in syntactic development. *Elementary English.* 43(7), pp. 732–739.

Hyltenstam, K. 1992. 'Non-Native Features of Near-Native Speakers: On the Ultimate Attainment of Childhood L2 Learners'. In R.J. Harris (ed.) *Cognitive Processing in Bilingual*. Amsterdam: Alselvier Science Publishers, B.V., pp. 351–368.

Ishikawa, S. 1995. Objective measurement of low-proficiency EFL narrative writing. *Journal of Second Language Writing*. 4(1), pp. 51–69.

Iwashita, N. 2006. Syntactic complexity measures and their relation to oral proficiency in Japanese as a foreign language. *Language Assessment Quarterly: An International Journal*. 3(2), pp. 151–169.

Iwashita, N. 2010. Features of oral proficiency in task performance by EFL and JFL learners. In *Selected proceedings of the 2008 second language research forum*, pp. 32–47.

James, C. 1998. *Error in Language Learning and Use: Exploring Error Analysis*. London: Longman.

Johansson, V. 2008. Lexical diversity and lexical density in speech and writing: a developmental perspective. *Working Papers in Linguistics*. *53*, pp. 61–79.

Johansson, S. 2009. 'Some Thoughts on Corpora and Second Language Acquisition'. In K. Ajmer (ed.) *Corpora and Language Teaching.* Amsterdam: Benjamins, pp. 33–46

Johnson, K. and Johnson, H. (eds). 1998. *Encyclopedic Dictionary of Applied Linguistics*. Oxford: Blackwell Publishers.

Johnson, M. 2004. *A Philosophy of Second Language Acquisition*. New York: Yale University Press.

Johnson, M.D., Mercado, L. and Acevedo, A. 2012. The effect of planning sub-processes on L2 writing fluency, grammatical complexity, and lexical complexity. *Journal of Second Language Writing*. 21(3), pp. 264–282.

Jokar, M. & Soyoof, A. 2014. The influence of written corrective feedback on two Iranian learners' grammatical accuracy. *Procedia-Social and Behavioral Sciences*. 98, pp. 799–805.

Jordan, G. 2004. *Theory Construction in Second Language Acquisition*. Amsterdam: Benjamins.

Jung, S. 2013. The effect of task repetition and corrective feedback in L2 writing: A pilot study. *MSU Working Papers in Second Language Studies*. 4(1), pp. 24–38.

Knoch, U. 2009. *Diagnostic Writing Assessment: The Development and Validation of a Rating Scale* (Vol. 17). Frankfurt: Peter Lang.

Kormos, J. 2011. Task complexity and linguistic and discourse features of narrative writing performance. *Journal of Second Language Writing*. 20(2), pp. 148–161.

Kuhi, D., Rasuli, M.A. & Deylami, Z. 2014. The effect of type of writing on accuracy, fluency and complexity across proficiency. *Procedia-Social and Behavioral Sciences*. 98, pp. 1036–1045.

Kuiken, F. & Vedder, I. 2008. Cognitive task complexity and written output in Italian and French as a foreign language. *Journal of Second Language Writing*. 17(1), pp. 48–60.

Kyle, K. & Crossley, S.A. 2015. Automatically assessing lexical sophistication: Indices, tools, findings, and application. *TESOL Quarterly*. 49(4), pp. 757–786.

Lambert, C. & Kormos, J. 2014. Complexity, accuracy, and fluency in task-based L2 research: Toward more developmentally based measures of second language acquisition. *Applied Linguistics*. 35(5), pp. 607–614.

Larsen-Freeman, D. 1978. An ESL index of development. *TESOL Quarterly*. 12(4), pp. 439–448.

Larsen-Freeman, D. 2006. The emergence of complexity, fluency, and accuracy in the oral and written production of five Chinese learners of English. *Applied linguistics*. 27(4), pp. 590–619.

Larsen-Freeman, D. 2009. Adjusting expectations: The study of complexity, accuracy, and fluency in second language acquisition. *Applied linguistics*. 30(4), pp. 579–589.

Laufer, B. & Nation, P. 1995. Vocabulary size and use: Lexical richness in L2 written production. *Applied linguistics*. 16(3), pp. 307–322.

Laufer, B., & Waldman, T. 2011. Verb–noun collocations in second language writing: A corpus analysis of learners' English. *Language Learning*. 61(2), pp. 647–672.

Lee, Y. J. 2006. The process-oriented ESL writing assessment: Promises and challenges. *Journal of Second Language Writing*. 15(4), pp. 307–330.

Leki, I., & Carson, J. 1994. Students' perceptions of EAP writing instruction and writing needs across the discipline. *TESOL Quarterly*. 28, pp. 81–101.

Lennon, P. 1991. Error: Some problems of definition, identification, and distinction. *Applied linguistics*. 12(2), pp. 180–196.

Lim, H.P. 1984. Measuring writing proficiency of college ESL students. *JALT JOURNAL*. 6, pp. 83–99.

Linnarud, M. (1986). *Lexis in Composition: A Performance Analysis of Swedish Learners' Written English*. Lund, Sweden: CWK Gleerup.

Lindqvist, C., Gudmundson, A. & Bardel, C., 2013. A new approach to measuring lexical sophistication in L2 oral production. *Eurosla Monographs Series 2*. pp 109–126.

Lintunen, P. & Mäkilä, M. 2014. Measuring syntactic complexity in spoken and written learner language: comparing the incomparable? *Research in Language*. 12(4), pp. 377–399.

Lo, J., & Hyland, F. 2007. Enhancing students' engagement and motivation in writing: The case of primary students in Hong Kong. *Journal of Second Language Writing*. 16(4), pp. 219–237.

Loban, W. 1963. *The language of elementary school children* (Research Rep. No.1). Champaign, IL: National Council of Teachers of English.

López, W.C. 2009. Error analysis in a learner corpus: what are the learners' strategies? In *A survey of corpus-based research [Recurso electrónico]*. pp. 675–690.

Lu, X. 2010. Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics*. 15(4), pp. 474–496.

Lu, X. 2011. A carpus-based evaluation of syntactic complexity measures as indices of college-level ESL writers' language development. *TESOl Quarterly*. 45(1), pp. 36–62.

Lu, X. 2012. The relationship of lexical richness to the quality of ESL learners' oral narratives. *The Modern Language Journal*. 96(2), pp. 190–208.

Lu, X. & Ai, H. 2015. Syntactic complexity in college-level English writing: Differences among writers with diverse L1 backgrounds. *Journal of Second Language Writing*. 29, pp. 16–27.

Lundstrom, K., & Baker, W. 2009. To give is better than to receive: The benefits of peer review to the reviewer's own writing. *Journal of Second Language Writing*. 18(1), pp. 30–43.

Macaro, E., & Masterman, L. 2006. Does intensive explicit grammar instruction make all the difference? *Language Teaching Research*. 10(3), pp. 297–327.

Malvern, D.D., & Richards, B.J. 1997. 'A New Measure of Lexical Diversity'. In A. Ryan & A. Wray (eds). *Evolving Models of Language*. Clevedon: Multilingual Matters, pp. 58-71.

Malvern, D., Richards, B., Chipere, N., & Dura´n, P. 2004. *Lexical Diversity and Language Development: Quantification and Assessment*. Houndmills, England: Palgrave MacMillan.

Marzban, A. & Arabahmadi, S. 2013. The effect of written corrective feedback on Iranian EFL students' writing. *Procedia-Social and Behavioral Sciences* 83, pp. 1000–1005.

McCarthy, S.J., Guo, Y.-H., & Cummins, S. 2005. Understanding changes in elementary Mandarin students' L1 and L2 writing. *Journal of Second Language Writing*. 14(2), pp. 71–104.

McCarthy, P.M. & Jarvis, S. 2007. *Vocd*: A theoretical and empirical evaluation. *Language Testing*. 24(4), pp. 459–488.

McClure, E. 1991. A comparison of lexical strategies in Ll and L2 written English narratives [Monograph]. *Pragmatics and Language Learning*. 2, pp. 141–154.

McCarthy, P.M. 2005. *An assessment of the range and usefulness of lexical diversity measures and the potential of the measure of textual, lexical diversity (MTLD)*. Unpublished PhD dissertation, the University of Memphis.

McKee, G., Malvern, D., & Richards, B. 2000. Measuring vocabulary diversity using dedicated software. *Literary and Linguistic Computing*. 15, pp. 323-337.

Meunier, F. 1998. 'Computer Tools for the Analysis of Learner Corpora'. In S. Granger (ed.) *Learner English on Computer*. New York: Addison Wesley Longman, pp. 19–37.

Meunier, F. 2012. Formulaic language and language teaching. *Annual Review of Applied Linguistics*. 32, pp. 111–129.

Miller, J. 2005. Most of ESL students have trouble with the articles. *International Education Journal*. 5(5), pp. 80–88.

Milton, J. 2009. *Measuring Second Language Vocabulary Acquisition*. Bristol: Multilingual Matters

Neary-Sundquist, C. 2014. 'The Development of Complexity in a Learner Corpus of German'. In J. Connor-Linton & L.W. Amoroso (eds). *Measured Language: Quantitative Approaches to Acquisition, Assessment, and Variation*. Washington D.C.: Georgetown University Press, pp. 213–227.

Nemser, W. 1971. Approximate systems of foreign language learners. *International Review of Applied Linguistics.* 9(2), pp. 115–123.

Nesselhauf, N. 2004. *Collocations in a Learner Corpus*. Amsterdam: Benjamins.

Ney, J. 1966. Review of grammatical structures written at three grade levels *Language Learning*. 16, pp. 230–235.

Nihalani, N.K. 1981. The quest for the L2 Index of Development. *RELC Journal*. 12(2), pp. 50–56.

Norris, J.M. & Ortega, L. 2009. Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics*.30(4), pp. 555–578.

O'Donnell, A.M., Reeve, J., & Smith, J.K. 2012. *Educational Psychology: Reflection for Action*. 3rd edn. NJ: John Wiley.

Ojima, M. 2006. Concept mapping as pre-task planning: A case study of three Japanese ESL writers. *System*. 34(4), pp. 566–585.

O'Loughlin, K. 1995. Lexical density in candidate output of direct and semi-direct versions of an oral proficiency test. *Language Testing*. 12, pp. 217–237.

Ong, J. & Zhang, L.J. 2010. Effects of task complexity on the fluency and lexical complexity in EFL students' argumentative writing. *Journal of Second Language Writing*. 19(4), pp. 218–233.

Ortega, L. 2003. Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college- level L2 writing. *Applied linguistics*. 24(4), pp. 492–518.

Pallotti, G. 2009. CAF: Defining, refining and differentiating constructs. *Applied Linguistics*. 30(4), pp. 590–601.

Paquot, M. 2008. 'Exeplification in Learner Writing: A Cross Linguistic Perspective'. In S. Granger &, F. Meunier (eds). *Phraseology in Foreign Language Learning and Teaching*. Amsterdam: John Benjamins, pp. 101–119.

Paquot, M. 2013. Lexical bundles and L1 transfer effects. *International Journal of Corpus Linguistics*. 18(3), pp. 391–417.

Polio, C.G. 1997. Measures of linguistic accuracy in second language writing research. *Language learning*. 47(1), pp. 101–143.

Polio, C., Fleck, C., & Leder, N. 1996 . *Second language essay revision and linguistic accuracy*Paper presented at the meeting for the Canadian Association for Applied Linguistics, London, Ontario.

Polio, C. & Shea, M.C. 2014. An investigation into current measures of linguistic accuracy in second language writing research. *Journal of Second Language Writing*. 26, pp. 10–27.

Quirk, R. & Greenbaum, S. 1973. *A University Grammar of English*. Harlow: Longman.

Rahimpour, M. & Hosseini, P. 2010. The impact of task complexity on L2 learners' written narratives. *English Language Teaching*. 3(3), pp. 198–205.

Read, J. 2000. *Assessing Vocabulary*. Oxford: Oxford University Press.

Rezazadeh, M., Tavakoli, M. & Rasekh, A.E. 2011. The Role of task type in foreign language written production: Focusing on fluency, complexity, and accuracy. *International Education Studies*. 4(2), pp. 169–176.

Riazantseva, A. 2012. Outcome measure of L2 writing as a mediator of the effects of corrective feedback on students' ability to write accurately. *System*. 40(3), pp. 421–430.

Robinson, P. 2001. Task complexity, task difficulty, and task production: Exploring interactions in a componential framework. *Applied linguistics*. 22(1), pp. 27–57.

Robinson, P. 2005. Cognitive complexity and task sequencing: Studies in a componential framework for second language task design. *IRAL-International Review of Applied Linguistics in Language Teaching*. 43(1), pp. 1–32.

Ruegg, R., Fritz, E., & Holland, J. 2011. Rater sensitivity to qualities of lexis in writing. *TESOL Quarterly*. 45(1), pp. 63–80.

Ruiz-Funes, M. 2015. Exploring the potential of second/foreign language writing for language learning: The effects of task factors and learner variables. *Journal of Second Language Writing*. 28, pp. 1–19.

Salimi, A., Dadaspour, S. & Asadollahfam, H. 2011. The effect of task complexity on EFL learners' written performance. *Procedia-Social and Behavioral Sciences*. 29, pp. 1390–1399.

Salimi, A., Alavinia, P. & Hosseini, P. 2012. The effect of strategic planning time and task complexity on L2 written accuracy. *Theory and Practice in Language Studies*. 2(11), pp. 2398–2406.

Salimi, A. & Dadashpour, S. 2012. Task complexity and language production dilemmas (Robinson's Cognition Hypothesis vs. Skehan's Trade-off Model). *Procedia-Social and Behavioral Sciences* 46, pp. 643–652.

Salimi, A. & Valizadeh, M. 2015. The effect of coded and uncoded written corrective feedback on the accuracy of learners writing in pre-intermediate level. *International Journal of Applied Linguistics and English Literature*. 4(3), pp. 116–122.

Salimi, A., Khayef, E. & Amini, D. 2015. The effect of strategic and online planning on accuracy of l1 learners' written production. *International Journal of Applied Linguistics and English Literature*. 4(6), pp. 148–153.

Sample, E. & Michel, M. 2014. An exploratory study into trade-off effects of complexity, accuracy, and fluency on young learners' oral task repetition. *TESL Canada Journal*. 31(8), pp. 23–46.

Sasaki, M. (2000). Toward an empirical model of EFL writing processes: An exploratory study. *Journal of Second Language Writing*. 9(3), pp. 259–291.

Sato, R. 2012. The effects of written feedback in the form of recasts. *The Journal of Asia TEFL*. 9(4), pp. 27–50.

Saville-Troike, M. 2012. *Introducing Second Language Acquisition*. Cambridge: Cambridge University Press.

Schachter, J. 1974. An error in error analysis. *Language Learning*. 24(2), pp. 205–214.

Schachter, J. & Celce-Murcia, M. 1977. Some reservations concerning error analysis. *Tesol Quarterly*. 11(4), pp. 441–451.

Schleppegrell, M.J. 1996. Conjunction in spoken English and ESL writing. *Applied Linguistics*. 17(3), pp. 271–285.

Scott, M.S. & Tucker, G. R. 1974. Error analysis and English- language strategies of Arab students. *Language Learning*. 24(1), pp. 69–97.

Selinker, L. 1972. Interlanguage. *IRAL-International Review of Applied Linguistics in Language Teaching*. 10(3), pp. 209–231.

Seyyedi, K., Ismail, S.A.M.M., Orang, M. & Nejad, M.S. 2013. The Effect of Pre-Task Planning Time on L2 Learners' Narrative Writing Performance. *English Language Teaching*. 6(12), pp. 1–9.

Sheen, Y., Wright, D., & Moldawa, A. 2009. Differential effects of focused and unfocused written correction on the accurate use of grammatical forms by adult ESL learners. *System*. 37(4), pp. 556–569.

Shintani, N., Ellis, R. & Suzuki, W. 2013. Effects of written feedback and revision on learners' accuracy in using two English grammatical structures. *Language Learning*. 64(1), pp. 103–131.

Shintani, N. & Ellis, R. 2015. Does language analytical ability mediate the effect of written feedback on grammatical accuracy in second language writing? *System*. 49, pp. 110–119.

Sinclair, J. 1991. *Corpus Concordance Collocation*. Oxford: Oxford University Press.

Sinclair, J. 1996. *Preliminary recommendations on corpus typology*. [Online]. Date accessed (28, Feb. 2016). EAGLES Document TCWG-CTYP/P. Available from http://www. ilc. pi. cnr. it/EAGLES/corpustyp/corpustyp. html.

Sinclair, J. 2005. 'Corpus and Text—Basic Principles'. In M. Wynne (ed.) *Developing Linguistic Corpora: A Guide to Good Practice*, Oxford: Oxbow Books, pp. 1–16.

Šišková, Z. 2012. Lexical richness in EFL students' narratives. *Language Studies Working Papers*. 4, pp. 26–36.

Skehan, P. 1996. 'Second Language Acquisition and Task‐ Based Instruction'. In J. Willis, & D. Willis (eds). *Challenge and Change in Language Teaching*. Oxford: Heinemann, pp. 17–30.

Skehan, P. 1998. *A Cognitive Approach to Language Learning*. Oxford: Oxford University Press.

Skehan, P. 2003. Task-based instruction. *Language teaching*. 36(01), pp. 1–14.

Skehan, P. 2009. Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics*. 30(4), pp. 510–532.

Skehan, P. and Foster, P. 1997. Task type and task processing conditions as influences on foreign language performance. *Language teaching research*. 1(3), pp. 185–211.

Soleimani, M., Modirkhamene, S. & Sadeghi, K. 2015. Peer-mediated vs. individual writing: measuring fluency, complexity, and accuracy in writing. *Innovation in Language Learning and Teaching*. DOI: 10.1080/17501229.2015.1043915, pp. 1–15.

Spoelman, M., & Verspoor, M. 2010. Dynamic patterns in development of accuracy and complexity: A longitudinal case study in the acquisition of Finnish. *Applied Linguistics*. 31(4), pp. 532–533.

Stevenson, M., Schoonen, R., & de Glopper, K. 2006. Revising in two languages: A multi-dimensional comparison of online writing revisions in L1 and FL. *Journal of Second Language Writing*. 15(3), pp. 201–233.

Storch, N. (2005). Collaborative writing: Product, process, and students' reflections. *Journal of Second Language Writing*. 14(3), pp. 153–173.

Storch, N. 2009. The impact of studying in a second language (L2) medium university on the development of L2 writing. *Journal of Second Language Writing*. 18(2), pp. 103–118.

Tai, H.Y. 2015. Writing development in syntactic complexity, accuracy and fluency in a content and language integrated learning class. *International Journal of Language and Linguistics.* 2(3), pp. 149–156.

Tapia, E. 1993. *Cognitive demand as a factor in interlanguage syntax. A study in topics and texts.* Unpublished dissertation, Indiana University.

Templin, M. 1957. *Certain Language Skills in Children: Their Development and Interrelationships*. Minneapolis: The University of Minnesota Press.

Thordardottir, E., & Ellis Weismer, S. 2001. High- frequency verbs and verb diversity in the spontaneous speech of school-age children with specific language impairment. *International Journal of Language and Communication Disorders*. 36, pp. 221–244.

Towell, R. 2012. 'Complexity, Accuracy and Fluency from the Perspective of Psycholinguistics Second Language Acquisition Research '. In A. Housen, F., Kuiken, F. & I., Vedder (eds). *Dimensions of L2 Performance and Proficiency Complexity, Accuracy and Fluency in SLA*. Philadelphia: John Benjamins, pp. 47–70.

Trawinski, M. (2005). *An outline of second language acquisition theories.* Krakow: Academic Pedagogic published journal and conference papers and presented at national and international levels.

Truscott, J., & Hsu, A.Y. 2008. Error correction, revision, and learning. *Journal of Second Language Writing*. 17(4), pp. 292–305.

Tweedie, F.J., & Baayen, R.H. 1998. How variable may a constant be? Measures in lexical richness in perspective. *Computers and the Humanities.* 32, pp. 323–352.

Ure, J. 1971. 'Lexical density: A computational Technique and Some Findings'. In M. Coultard (ed.) *Talking about Text*. Birmingham, England: English Language Research, University of Birmingham, pp. 27–48.

Vann, R.J. 1978. *A study of the oral and written English of adult Arabic speakers*. Unpublished Ph.D. dissertation, Indiana University.

Vesal, N.S., Vesal, M.S. & Tavakoli, M. 2015. The effect of task type on complexity, accuracy, and fluency of Iranian EFL candidates' oral production: IELTS interview test in focus. *Cumhuriyet Science Journal* 36(3), pp. 1154–1168.

Vyatkina, N. 2012. The development of second language writing complexity in groups and individuals: A longitudinal learner corpus study. *The Modern Language Journal*. 96(4), pp. 576–598.

Wang, X. 2014. The relationship between lexical diversity and EFL writing proficiency. *University of Sydney Papers in TESOL*. 9, pp. 65–88.

Wang, Y. 2008. Influence of planning on students' language performance in task-based language teaching. *English language teaching*. 1(1), pp. 83–86.

Wardhaugh, R. 1970. The contrastive analysis hypothesis. *TESOL Quarterly*. 4(2) pp. 123–130.

Way, D., Joiner, E., & Seaman, M. (2000). Writing in the secondary foreign language classroom: The effects of prompts and task on novice learners of French. *Modern Language Journal*. 84(2), pp. 171–184.

Wigglesworth, G., & Foster, P., 2008, April. Measuring accuracy in second language performance, Paper presented at the TESOL Convention, New York City.

Wolfe-Quintero, K., Inagaki, S., & Kim, H.-Y. 1998. *Second Language Development in Writing: Measures of Fluency, Accuracy, and Complexity*. USA: University of Hawai'i Press.

Yuan, F. & Ellis, R. 2003. The effects of pre- task planning and on- Line planning on fluency, complexity and accuracy in L2 monologic oral production. *Applied linguistics*. 24(1), pp. 1–27.

Yang, W., & Sun, Y. 2015. Dynamic development of complexity, accuracy and fluency in multilingual learners' L1, L2 and L3 writing. *Theory and Practice in Language Studies*. 5(2), pp. 298–308.